实用偏微分方程数值解法

徐 长 发

华中理工大学出版社

·研究生用书·

实用偏微分方程数偏解法

徐长发

责任编辑　李立鹏

*

华中理工大学出版社出版发行

（武昌喻家山）

新华书店湖北发行所经销

华中理工大学出版社印刷厂印刷

*

开本：850×1168　1/32　印张：16.125　插页：2　字数：390 000

1990年12月第1版　　1990年12月第1次印刷

印数：1—1 000

ISBN7--5609—0511—0/O·75

定价：3.48元

内 容 简 介

本书共三篇。第一篇讨论了抛物型和双曲型方程的差分方法，介绍了各种实用的差分格式及其稳定性分析。第二篇讨论了椭圆型方程的有限元方法，清晰揭示了基本思想、应用技巧、通用程序设计和基本理论问题。篇三篇讨论解常微分方程的高效率高精度方法。

本书题材新颖，利于实用，内容深入浅出，便于自学，内容丰富，便于选用，或侧重于算法与应用，或算法与分析并重。

本书可作为高等院校理工科各专业高年级学生和研究生的教材，也可供有关科研和工程技术人员参考。

ABSTRACT

This book consists of three parts. The difference method for parabolic and hyperbolic equations is discussed in the first part, where there are many practical schemes, design ideas, and their numerical atability-analyses. The finite element method for elliptic equations is given in the second part, where the fundamental theories and the applied techniques etc. are discussed. The varied new numerical methods for discrete systems are in the third part, where these highly accurate and efficent algorithms and the applied techniques are expounded.

The content of this book is extensive, practi[illegible] and easy to comprehend.

This book can suit the needs of graduates of science end engineering, and is also useful for engineers end researchers.

[illegible]

“研究生用书”总序

研究生教材建设是提高研究生教学质量的重要环节，是具有战略性的基本建设。各门课程必须有高质量的教材，才能使学生通过学习掌握各门学科的坚实的基础理论和系统的专门知识，为从事科学研究工作或独立担负专门技术工作打下良好的基础。

我校各专业自1978年招收研究生以来，组织了一批学术水平较高、教学经验丰富的教师，先后编写了公共课、学位课所需的多种教材和教学用书。有的教材和教学用书已正式出版发行，更多则采用讲义的形式逐年印发。这些讲义经过任课教师多年教学实践，不断修改、补充、完善，已达到出书的要求。因此，我校决定出版“研究生用书”，以满足本校各专业研究生教学需要，并与校外单位交流，征求有关专家学者和读者的意见，以促进我校研究生教材建设工作，提高教学质量。

“研究生用书”以公共课和若干门学位课教材为主，还有教学参考书和学术专著，涉及的面较广，数量较多，准备在今后数年内分批出版。编写“研究生用书”总的要求是从研究生的教学需要出发，根据各门课程在教学过程中的地位和作用，在内容上求新、求深、求精，每本教材均应包括本门课程的基本内容，使学生能掌握必需的基础理论和专门知识；学位课教材还应接触该学科的发展前沿，反映国内外的最新研究成果，以适应目前科学技术

知识更新很快的形势；学术专著则应充分反映作者的科研硕果和学术水平，阐述自己的学术见解。在结构和阐述方法上，应条理清楚，论证严谨，文字简炼，符合人们的认识规律。总之，要力求使“研究生用书”具备科学性、系统性和先进性。

我们的主观愿望虽然希望“研究生用书”的质量尽可能高一些，但由于研究生的培养工作为时尚短，水平和经验都不够，其中缺点、错误在所难免，尚望校内外专家学者及读者不吝指教，我们将非常感谢。

华中理工大学研究生院院长

陈 珽

1989.11.

前　言

随着科学技术的飞速发展，偏微分方程的数值解法得到了深入的研究和广泛的应用。应用数学专业的学生和研究生，计算数学专业的学生，理工科各专业的研究生，以及有关的科研、工程技术人员等都迫切需要学习和掌握偏微分方程数值处理的基本方法、基本理论和应用技能。为此目的，作者在原讲义多次使用并广泛听取有关方面意见的基础上，经反复修改后写成了这本书。

众所周知，偏微分方程可根据它的数学特征分为三大类型，即抛物型、双曲型和椭圆型。这三类偏微分方程描述了不同本质的物理现象，其应用是极其广泛的。对于在理论研究和实际应用问题中提出的许多偏微分方程，由于其边界和边界条件复杂等原因，寻求解的解析表达式相当困难，有时甚至是不可能的，所以必须利用计算机研究微分方程的数值解。简言之，这种研究的任务在实用中主要表现于两个方面。一是关于用有效的数值方法离散偏微分方程及其边界条件。对此，差分方法和有限元方法是目前被普遍认为行之有效的两类主要的数值方法。二是关于高效率高精度求解离散微分方程。对此，解同样的离散微分方程，采用好的算法与采用一般算法的计算效果往往相差很大，采用好的算法不但能使求解过程数值稳定、数值解的精度得到提高，而且能数十倍、数百倍地节省计算工作量。

本书共三篇。第一篇介绍解抛物型和双曲型方程初（边）值问题的差分方法；第二篇介绍解椭圆型方程边值问题的有限元方法；第三篇介绍解离散微分方程的高效率方法。这三篇构成一个整体，介绍了偏微分方程数值处理中的基本思想、有关理论、有效算法、程序设计技巧和数值例子等内容。本书章节与段落的标题鲜明，读者可根据自己的要求选用，或侧重于算法与应用，或

算法与分析并重，因此本书能适应不同层次教学和阅读的需要。

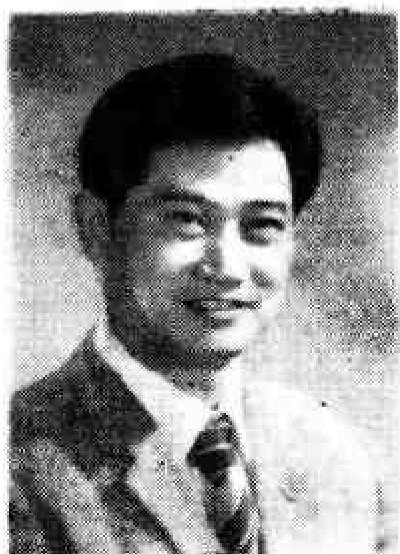
本书注意了以下四方面情况。第一，注重基本思想，不求面面俱到。例如差分方法和有限元方法，虽然这两种离散化方法对抛物型、双曲型和椭圆型这三类偏微分方程都可应用，但本书只就某种离散化方法在其应用最广泛的方面加以侧重介绍，而且非常强调学习和应用这些离散化方法的主要思想和基本方法，以便读者尽快掌握这些方法并自行扩大应用范围。第二，注意应用和实践。为此目的，我们以分析问题解决问题的方法为主导，且以算法为线索展开全书内容，书中除了指明具体算法的计算公式和实现过程外，还特别安排了有关程序设计的内容，列出了便于应用和参考的计算框图及算例数据分析等，这些都有利于提高读者实践能力和设计思想。第三，反映最新成果。特别是将一些最近研究成熟的既实用又有效的算法介绍给读者。第四，内容阐述方式力求数学化和现代化，并附有深入浅出的解释。这既有利于具备数值分析和数理方程初步知识的广大读者顺利自学，又有利于读者在数学上的提高，还便于读者学习和研究有关参考文献。

黄鸿生教授审阅了全书，还有不少同志为本书编写和出版给予过帮助，作者在此一并表示衷心的感谢。

由于作者水平有限，书中错误在所难免，敬请广大读者指正。

徐长发

于华中理工大学　1989年3月



作者简介

徐长发，1947年生，1970年毕业于华中工学院机械制造系，现任华中理工大学数学系副教授。曾于1979～1981年赴西德波恩大学进修计算数学。多年来一直从事偏微分方程数值解法的教学和研究工作，并在学术刊物上发表了论文十余篇。

● **研究生用书**

● 责任编辑：李立鹏

● 封面设计：俞缦丽　陈建纲

● 华中理工大学出版社

目录

第一篇　解抛物型和双曲型方程的差分方法

抛物型和双曲型方程描述了物质扩散和波动等不定常物理过程，这两类偏微分方程的定解问题在力学、热传导理论、燃烧理论、化学、空气动力学、电磁学和经济数学等方面都有着广泛的应用。这两类方程一般可转化为常微分方程(组)，它们在解的结构形式和解的稳定性研究方面都有共同之处，因此本篇重点讨论具有代表性的初值问题的差分方法。

初值问题的差分方法是古老的，早在1928年，Courant、Friedrichs和Lewy在《论数学物理中的偏微分方程差分方法》一文中就作过较深入的讨论。自从电子计算机问世后，差分方法的研究和应用迅速地发展了，近几十年来新的思想和算法仍不断出现，就目前的研究和应用现状看，解初值问题的差分方法是普遍实用的。

本篇主要就线性方程的初值问题介绍各种有效实用的差分格式和构造这些格式的思想方法，同时也特别注重这些差分格式的稳定性分析。本篇所介绍的种种离散化方法中，有不少是具有启发性的。有的给出了构造高精度差分格式的方法，有的表明了保证差分格式数值稳定的技巧，有的兼顾了实际应用的适应性和计算效率，有的还揭示了不同类型偏微分方程之间的转换规律。读者若能将这些算法融汇贯通，无疑是极其有益的。

第一章　解抛物型方程的差分方法

本章主要讨论一般性的二阶线性抛物型方程，其定解问题假定是适定的，特别是重点讨论热传导方程这一模型问题的差分方法。本章先提出解抛物型方程的差分方法中必须讨论的有关理论问题及其研究的方法，再介绍各种差分格式的构造及其稳定性分析。本章内容不仅对解非线性抛物型方程差分方法有指导意义，而且也是学习第二章的基础。

§1　二阶线性抛物型方程的适定性及其解结构

考虑最简单的**抛物型方程**

$$Lu=\frac{\partial u}{\partial t}-a^2\frac{\partial^2 u}{\partial x^2}=f(x,t),\quad a>0,\quad (x,t)\in D,\qquad (1.1)$$

其中，D是 x-t 平面内的给定区域，它可以是有界区域，也可以是无界区域，L是**微分算子**。

方程（1.1）的**定解问题**根据**定解条件**可分为四种类型：

1. 初值问题(Cauchy问题)

在区域$D=\{(x,t)|x\in(-\infty,+\infty),\ t>0\}$内求方程(1.1)满足初始条件

$$u(x,0)=\varphi(x),\qquad x\in(-\infty,+\infty)\qquad (1.2)$$

的解$u(x,t)$。

2. 第一类初边值问题

在区域 $D=\{(x,t)|x\in(0,l),\ t\in(0,T)\}$ 内求方程（1.1）满足初始条件和第一类边值条件

$$\begin{cases}u(x,0)=\varphi(x), & x\in(0,l),\\ u(0,t)=\alpha_1(t), \quad u(l,t)=\alpha_2(t), & t\in(0,T)\end{cases} \tag{1.3}$$

的解$u(x,t)$。

3．第二类初边值问题

在区域$D=\{(x,t)|x\in(0,l),\ t\in(0,T)\}$内求方程（1.1）满足初始条件和第二类边值条件

$$\begin{cases}u(x,0)=\varphi(x), & x\in(0,l),\\ u_x(0,t)=\beta_1(t),\ u_x(l,t)=\beta_2(t), & t\in(0,T)\end{cases} \tag{1.4}$$

的解$u(x,t)$。

4．第三类初边值问题

在区域$D=\{(x,t)|x\in(0,l),\ t\in(0,T)\}$内求方程（1.1）满足初始条件和第三类边值条件

$$\begin{cases}u(x,0)=\varphi(x), & x\in(0,l),\\ (u-\delta_1(t)u_x)|_{x=0}=\gamma_1(t), & \\ (u-\delta_2(t)u_x)|_{x=l}=\gamma_2(t), & t\in(0,T)\end{cases} \tag{1.5}$$

的解$u(x,t)$。

在抛物型方程的研究中，定解问题的适定性概念是重要的。如果定解问题的解存在唯一而且稳定，则称定解问题是适定的。

某个物理过程在一定条件下总是具有唯一确定的状态，因此正确描述这种物理过程的定解问题的解应是存在唯一的。但是用偏微分方程对此作出描述时，总要经过一些近似处理（例如要舍弃一些因素）并提出一些附加要求，我们只能说定解问题近似地反映了自然现象中某个物理过程，而不能说两者完全等同。因此，完全有必要从量的角度来研究解的存在性和唯一性。只有弄清这一问题后，才能判断对于归结出来的定解问题，所给定的条件是否足以保证解的唯一性。如果条件不完备，则还要找新的条件。

此外，出现在定解条件中的函数值通常是由实验测量确定的，一般只是近似的。在这种定解条件下求得的解，只有在它和准确定解条件下确定的解相差很小时，才有实际意义。如果定解条件的微小变化引起的解在定义域中的变化也是微小的，则称定解问题是**稳定的**，或称解是连续依赖于定解条件的。

下面以定解问题（1.1）-（1.2）、（1.1）-（1.3）为例，列出数理方程研究中有关适定性的部分结论。

利用数理方程中求**基本解**的办法，可求得齐次初值问题

$$\begin{cases} \dfrac{\partial u}{\partial t} - a^2\dfrac{\partial^2 u}{\partial x^2} = 0, \\ u(x,0) = \varphi(x), \end{cases} \quad a>0,\ t>0,\ x\in(-\infty,+\infty) \tag{1.6}$$

的适定解为

$$u(x,t) = \int_{-\infty}^{+\infty} G_1(x,\xi,t)\varphi(\xi)\,d\xi, \tag{1.7}$$

其中，

$$G_1(x,\xi,t) = \frac{1}{2a\sqrt{\pi t}}\exp[-(x-\xi)^2/4a^2t]$$

是齐次初值问题（1.6）的基本解。若$\varphi(x)$是有界连续函数，则可确定（1.7）是（1.6）的解；不仅如此，$u(x,t)$还具有较高的光滑性，它可以对自变量x和t微分任意多次。

对于非齐次初值问题

$$\begin{cases} \dfrac{\partial u}{\partial t} - a^2\dfrac{\partial^2 u}{\partial x^2} = f(x,t), \\ u(x,0) = \varphi(x), \end{cases} \quad a>0,\ t>0,\ x\in(-\infty,+\infty), \tag{1.8}$$

也可求得适定解为

$$u(x,t) = \int_{-\infty}^{+\infty} G_1(x,\xi,t)\varphi(\xi)\,d\xi + \int_0^t\int_{-\infty}^{+\infty} G_1(x,\xi,t-\tau)f(\xi,\tau)\,d\xi d\tau. \tag{1.9}$$

由（1.7）和（1.9）式可知，Cauchy 问题的解由两部分构成：一部分是初值的影响，它是由初值与基本解的卷积构成；另一部分是右端项的影响，它是由右端项与基本解的卷积构成。

对于方程

$$
\begin{cases}
\dfrac{\partial u}{\partial t}-a^2\dfrac{\partial^2 u}{\partial x^2}=0, & a>0,\ t>0,\ x\in(0,l),\\
u(x,0)=\varphi(x), & x\in(0,l),\\
u(0,t)=u(l,t)=0, & t\geqslant 0
\end{cases}
\tag{1.10}
$$

可求得适定解为

$$
u(x,t)=\int_0^l G_2(x,\xi,t)\varphi(\xi)d\xi, \tag{1.11}
$$

其中，

$$
G_2(x,\xi,t)=\frac{2}{l}\sum_{n=1}^{\infty}\exp[-(n\pi/l)^2a^2t]\sin\frac{n\pi}{l}x\sin\frac{n\pi}{l}\xi
$$

是定解问题（1.10）所对应的零初值情形的基本解。

对于方程

$$
\begin{cases}
\dfrac{\partial u}{\partial t}-a^2\dfrac{\partial^2 u}{\partial x^2}=f(x,t), & a>0,\ t>0,\ x\in(0,l),\\
u(x,0)=0, & x\in(0,l),\\
u(0,t)=u(l,t)=0, & t\geqslant 0
\end{cases}
\tag{1.12}
$$

也可求得适定解为

$$
u(x,t)=\int_0^t\int_0^l G_2(x,\xi,t-\tau)\cdot f(\xi,\tau)d\xi d\tau. \tag{1.13}
$$

对于方程

$$
\begin{cases}
\dfrac{\partial u}{\partial t}-a^2\dfrac{\partial^2 u}{\partial x^2}=f(x,t), & a>0,\ t>0,\ x\in(0,l),\\
u(x,0)=\varphi(x), & x\in(0,l),\\
u(0,t)=u(l,t)=0, & t\geqslant 0,
\end{cases}
\tag{1.14}
$$

只要令$u(x,t)=u_1(x,t)+u_2(x,t)$，其中$u_1(x,t)$和$u_2(x,t)$分

别是定解问题（1.10）和(1.12)的解，则可求得定解问题(1.14)的适定解为

$$u(x,t)=\int_0^l G_2(x,\xi,t)\varphi(\xi)d\xi$$

$$+\int_0^t\int_0^l G_2(x,\xi,t-\tau)f(\xi,\tau)d\xi d\tau. \qquad (1.15)$$

由（1.10）～(1.15）式可知，初边值混合问题与初值问题在解的结构方面有着共同的特点．另外，非齐次边界条件的混合边值问题可以化为齐次边界条件的混合边值问题，又可进一步作周期延拓．所以，对于抛物型方程解的结构和稳定性研究方面，研究初值问题是有代表性的．

下面以初值问题（1.8）为例，分析初值稳定与右端稳定的关系．

在抛物型定解问题中，若对于$\|\varphi(x)\|<+\infty$和给定的t，有

$$\|u(x,t)\|\leqslant K\|\varphi(x)\|, \qquad K=\text{const}, \qquad (1.16)$$

则称初值问题的解是关于$\|\cdot\|$**初值稳定的**．（1.16）式中的$\|\cdot\|$可定义为L_2模或最大模，即

$$\|u(x,t)\|_2^2=\int_{-\infty}^{+\infty}|u(x,t)|^2dx,$$

$$\|u(x,t)\|_\infty=\max_x|u(x,t)|.$$

由初值稳定的定义可知，初值问题（1.8）的解关于$\|\cdot\|$初值稳定的充要条件是基本解$G_1(x,t)$关于$\|\cdot\|$一致有界，即

$$\|G_1(x,t)\|\leqslant K, \qquad \forall(x,t)\in D. \qquad (1.17)$$

初值稳定的意义是明显的．设定解问题（1.8）的右端项$f(x,t)$无误差，仅是初值$\varphi(x)$有误差，设初值$\varphi(x)$和$\tilde{\varphi}(x)$所对应的初值问题的解是$u(x,t)$和$\tilde{u}(x,t)$，那么利用（1.9）、(1.16）和（1.17）式，就有

$$\|\tilde{u}(x,t)-u(x,t)\|=\|\int_{-\infty}^{+\infty}G_1(x,\xi,t)\tilde{\varphi}(\xi)d\xi$$

$$+\int_0^t\int_{-\infty}^{+\infty}G_1(x,\xi,t-\tau)f(\xi,\tau)d\xi d\tau-\int_{-\infty}^{+\infty}G_1(x,\xi,t)\varphi(\xi)d\xi$$

$$-\int_0^t\int_{-\infty}^{+\infty}G_1(x,\xi,t-\tau)f(\xi,\tau)d\xi d\tau\Big\|\leqslant K\|\tilde{\varphi}(x)-\varphi(x)\|.$$

这就是说，解的误差可由初值的误差来控制，或者说解是连续依赖于初值$\varphi(x)$的。

由上式可知，研究齐次方程和非齐次方程的解的初值稳定性是等效的。

在抛物型定解问题中，如果$\|f(x,t)\|<+\infty$且

$$\|u(x,t)\|\leqslant M\|f(x,t)\|,\qquad M=\mathrm{const},$$

则称初值问题的解是关于$\|\cdot\|$**右端稳定的**。

右端稳定的意义也是明显的。设定解问题（1.8）的初始值$\varphi(x)$无误差，仅是方程的右端项$f(x,t)$存在误差，若$\tilde{f}(x,t)$和$f(x,t)$所对应的初值问题的解是$\tilde{u}(x,t)$和$u(x,t)$，则有

$$\|\tilde{u}(x,t)-u(x,t)\|\leqslant M\|\tilde{f}(x,t)-f(x,t)\|.$$

这就是说，初值问题解的误差可由右端项的误差来控制，或解是连续依赖于右端项$f(x,t)$的。

由初值稳定和右端稳定的定义可推知：若初值问题的解关于模$\|\cdot\|$初值稳定，则必关于模$\|\cdot\|$右端稳定。

事实上，若定解问题（1.8）的解是关于$\|\cdot\|$初值稳定的，则有

$$\|G_1(x,t)\|<K,\qquad \forall(x,t)\in D.$$

若再设问题（1.8）中对应$\tilde{f}$和f的解分别为$\tilde{u}$和u，那么

$$\begin{aligned}\|\tilde{u}(x,t)-u(x,t)\|=\Big\|&\int_{-\infty}^{+\infty}G_1(x,\xi,t)\varphi(\xi)d\xi\\&+\int_0^t\int_{-\infty}^{+\infty}G_1(x,\xi,t-\tau)\tilde{f}(\xi,\tau)d\xi d\tau\\&-\int_{-\infty}^{+\infty}G_1(x,\xi,t)\varphi(\xi)d\xi\\&-\int_0^t\int_{-\infty}^{+\infty}G_1(x,\xi,t-\tau)f(\xi,\tau)d\xi d\tau\Big\|\end{aligned}$$

$$\leqslant M\|\tilde{f}(x,t)-f(x,t)\|.$$

综上所述，研究齐次方程初值问题的解的初值稳定性是有代表性的。

关于一般二阶线性抛物型初值问题的适定性和解结构的讨论有如下结论。

设 $x=(x_1,x_2,\cdots,x_n)\in E_n$，如果$E_n$是有界闭区域，二阶线性抛物型初值问题

$$\begin{cases} Lu=\sum\limits_{i,j=1}^{n}a_{ij}(x,t)\dfrac{\partial^2 u}{\partial x_i\partial x_j}+\sum\limits_{i=1}^{n}b_i(x,t)\dfrac{\partial u}{\partial x_i}+c(x,t)u \\ \qquad -\dfrac{\partial u}{\partial t}=f(x,t), \quad x\in E_n,\ t>0, \\ u(x,0)=\varphi(x),\ x\in E. \end{cases}$$

满足下列假设：

(1) 对$\forall(x_i,t_j)\in E_n\times(0,t)$，有$a_{ij},b_i,c,f$连续；

(2) 记$\beta=(\beta_1,\beta_2,\cdots,\beta_n)$，有

$$\sum_{i,j=1}^{n}a_{ij}(x,t)\beta_i\beta_j>0,\qquad \forall(x,t)\in E_n\times(0,t);$$

(3) $\varphi(x)$在E_n上连续；

则在$E_n\times[0,t]$内，此初值问题是适定的，其解为

$$u(x,t)=\int_{E_n}G(x,\xi,t)\varphi(\xi)d\xi+\int_0^t\int_{E_n}G(x,\xi,t-\tau)f(\xi,\tau)d\xi d\tau$$

其中，$G(x,\xi,t)$ 是此问题的基本解。

§2 古典差分格式

2.1 差分方法的基本步骤

考虑一维热传导方程的第一类初边值问题

$$
\begin{cases}
Lu = \dfrac{\partial u}{\partial t} - a^2 \dfrac{\partial^2 u}{\partial x^2} = f(x,t), & (x,t)\in D, \\
u(x,0) = \varphi(x), & x \in (0,1), \\
u(0,t) = u(1,t) = 0, & t \in [0,T],
\end{cases}
\tag{1.18}
$$

其中$\overline{D}=\{(x,t)\,|\,x\in[0,1],\ t\in[0,T]\}$，$a$为正常数。

用差分方法求解（1.18）问题的基本步骤如下：

第一步，定解区域离散化。用平行直线族$x_j=jh$，$t_k=k\tau$把区域$\overline{D}$分成若干个小矩形。位于区域D内部的网格节点(x_j,t_k)（简记为(j,k)）称为**内节点**，所有内节点用D_h表示。位于边界上的节点称作**界点**，所有界点记为∂D_h。h和τ分别称作**空间步长**和**时间步长**。如果$[0,1]$被分成N等分，$j=0,1,\cdots,N$，则$h=1/N$；如果$[0,T]$被分成M等分，$k=0,1,\cdots,M$，则$\tau=1/M$。参见图1.1。

第二步，定解问题离散化。由Taylor展式可知，在节点(j,k)处微商和差商有下述关系：

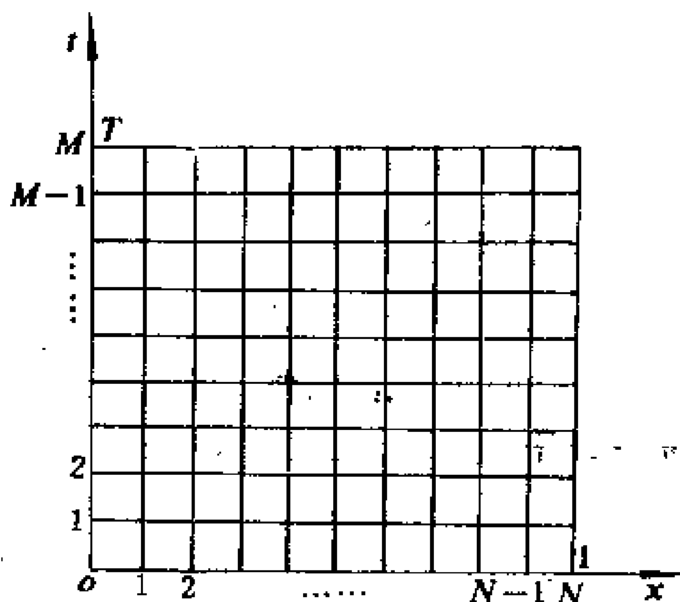


图 1.1

$$\begin{cases} \dfrac{u(x_j,t_{k+1})-u(x_j,t_k)}{\tau}=\left[\dfrac{\partial u}{\partial t}\right]_j^k+O(\tau), \\ \dfrac{u(x_{j-1},t_k)-2u(x_j,t_k)+u(x_{j+1},t_k)}{h^2}=\left[\dfrac{\partial^2 u}{\partial x^2}\right]_j^k+O(h^2). \end{cases} \tag{1.19}$$

若记$[u]_j^k=u(x_j,t_k)$，$f_j^k=f(x_j,t_k)$，u_j^k是$u(x_j,t_k)$的近似值，则用差商近似代替（1.18）式中的微商后，即得相应的**差分方程**，亦称**差分格式**：

$$L_h u_j^k=L_h[u]_j^k\equiv\frac{u_j^{k+1}-u_j^k}{\tau}-a^2\frac{u_{j-1}^k-2u_j^k+u_{j+1}^k}{h^2}=f_j^k, \tag{1.20}$$

$$j=1,2,\cdots,N-1;\ \ k=0,1,2,\cdots,M-1.$$

差分方程（1.20）称为解热传导方程（1.18）的**古典显格式**，它是一个两层四点格式，它所用到的节点图式如图1.2所示。

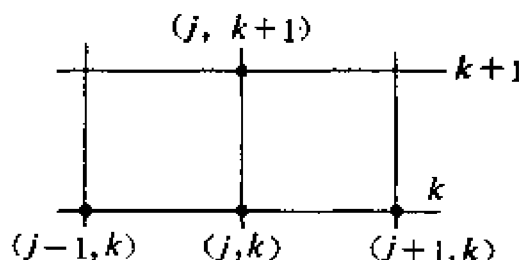


图 1.2

由（1.19）和（1.20）式，有

$$R_j^k:=L_h[u]_j^k-[Lu]_j^k=O(\tau+h^2),$$

称R_j^k为差分方程（1.20）的**截断误差**，这个误差量不仅反映了**差分算子**逼近**微分算子**的关系，也直接影响差分解和方程解之误差$|u_j^k-u(x_j,t_k)|$。

第三步，解差分方程。将定解条件也离散化，并引进“网比”记号$r=a^2\tau/h^2$，再将差分方程（1.20）改写成便于计算的形式，就有

$$
\begin{cases}
u_j^{k+1} = u_j^k + r[u_{j-1}^k - 2u_j^k + u_{j+1}^k] + \tau f_j^k \\
\qquad = ru_{j-1}^k + (1-2r)u_j^k + ru_{j+1}^k + \tau f_j^k, \\
u_j^0 = \varphi(x_j), \qquad u_0^k = u_N^k = 0, \\
r = a^2\tau/h^2, \qquad j = 1,2,\cdots,N-1, \qquad k = 1,2,\cdots,M-1.
\end{cases}
\tag{1.21}
$$

这样，利用$k=0$层上的数值，即可逐点算出 $k=1$ 层上的全部离散节点处解的近似值，利用 k 层上的节点值算出$k+1$层上的节点值。这样逐层逐点地显式计算数值解是非常方便的。

2.2 三个古典差分格式

古典显格式. 由（1.20）式可见，用向前差商代替时间方向的一阶偏导，用中心差商代替空间方向的二阶偏导，就得到截断误差为$O(\tau+h^2)$的古典显格式（1.20）和便于数值计算的(1.21)式。若引入向量$U^k=(u_1^k,\cdots,u_{N-1}^k)^T$，$F^k=(f_1^k,\cdots,f_{N-1}^k)^T$，就可把古典显格式写成矩阵形式：

$$
\begin{aligned}
U^{k+1} &= [(1-2r)I + rC]U^k + \tau F^k \\
&= AU^k + \tau F^k,
\end{aligned}
\tag{1.22}
$$

其中，

$$
A = \begin{pmatrix}
1-2r & r & & \\
r & 1-2r & r & \\
& \ddots & \ddots & \ddots \\
& & & r \\
& & r & 1-2r
\end{pmatrix}, \quad
C = \begin{pmatrix}
0 & 1 & & \\
1 & 0 & 1 & \\
& \ddots & \ddots & \ddots \\
& & & 1 \\
& & 1 & 0
\end{pmatrix},
$$

I为$N-1$阶单位矩阵。

古典隐格式. 仿照构造古典显格式中所采用的方法，即用差商直接代替微商的**直接差分化方法**，现用向后差商代替（1.18）式中一阶偏导，用二阶中心差商代替二阶偏导，就得到截断误差为$O(\tau+h^2)$的古典隐格式：

$$
\frac{1}{\tau}(u_j^k - u_j^{k-1}) - \frac{a^2}{h^2}(u_{j-1}^k - 2u_j^k + u_{j+1}^k) = f_j^k. \tag{1.23}
$$

这也是一个两层四点格式，它所涉及的节点如图1.3所示。

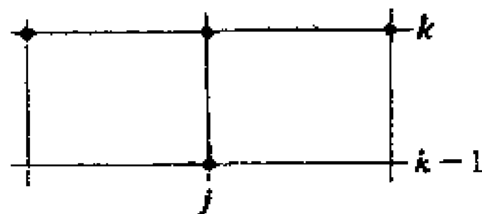


图 1.3

古典隐格式的截断误差 $R_j^k = O(\tau + h^2)$。

结合定解条件，就可将古典隐格式改写成便于数值计算的形式：

$$
\begin{cases}
u_j^k - r(u_{j-1}^k - 2u_j^k + u_{j+1}^k) = u_j^{k-1} + \tau f_j^k, \\
u_j^0 = \varphi(x_j), \qquad u_0^k = u_N^k = 0, \\
r = a^2\tau/h^2, \quad j = 1,2,\cdots,N-1, \quad k = 1,2,\cdots,M.
\end{cases}
\tag{1.24}
$$

古典隐格式的矩阵形式为

$$
AU^{k+1} = U^k + \tau F^{k+1}, \tag{1.25}
$$

其中，

$$
A = [(1+2r)I - rC] = \begin{pmatrix}
1+2r & -r & & & \\
-r & 1+2r & & & \\
 & \ddots & \ddots & \ddots & \\
 & & & & -r \\
 & & & -r & 1+2r
\end{pmatrix}.
$$

由此可见，差分方程组（1.24）的系数矩阵是一个强对角占优矩阵，故（1.24）式唯一可解。另外，它又是一个三对角矩阵，故差分方程组（1.24）可用追赶法求解。值得指出的是，对任一时间层 k，只有通过求解一个 $N-1$ 阶线性代数方程组，才能求得这一时间层上的古典隐格式差分解。显然隐式求解比显式求解复杂，但由下节的讨论可知，隐格式的稳定性较好。

Richardson 格式. 为了提高截断误差的阶，用一阶中心差商代替时间方向的一阶偏导，用二阶中心差商代替空间方向的二阶偏导，可得Richardson格式：

$$\frac{1}{2\tau}(u_j^{k+1}-u_j^{k-1})-\frac{a^2}{h^2}(u_{j-1}^k-2u_j^k+u_{j+1}^k)=f_j^k. \quad (1.26)$$

这是一个三层五点格式，它所涉及的节点如图1.4所示。

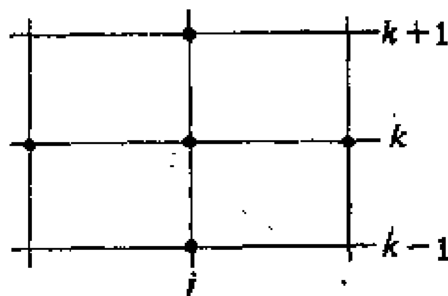


图 1.4

Richardson 格式的截断误差 $R_j^k=O(\tau^2+h^2)$。

结合定解条件，可将 Richardson 格式改写成便于数值计算的形式：

$$\begin{cases} u_j^{k+1}=u_j^{k-1}+2r(u_{j-1}^k-2u_j^k+u_{j+1}^k)+2\tau f_j^k, \\ u_j^0=\varphi(x_j), \quad u_0^k=u_N^k=0, \\ r=a^2\tau/h^2, \ j=1,2,\cdots,N-1, \ k=1,2,\cdots,M-1. \end{cases} \quad (1.27)$$

Richardson 格式是三层显格式，只有同时利用第$k-1$层和第k层上的全部数值解，才能逐点计算出第$k+1$层上的数值解。为此，必须先利用某种格式(例如古典显格式)算出$k=1$层上的数值解，然后才能利用Richardson 格式逐层地进行显式计算。

Richardson格式的矩阵形式为

$$U^{k+1}=U^{k-1}+2r(C-2I)U^k+2\tau F^k. \quad (1.28)$$

2.3 解抛物型方程的差分方法中的两个基本问题

前面，我们已用直接差分化方法构造了三种差分格式，读者自然会问，这些格式是否都可用于实际计算呢？这就要对差分格式进行理论分析，考虑到实际计算时初始值有误差（例如初始条件可能被干扰，也可能实验值有误差等），那么实际计算值 $\tilde{u}_j^k$ 和差分方程准确解u_j^k之间，以及它们和微分方程解 $u(x_j,t_k)$ 之间都存在误差。显然，要求数值计算值逼近原方程的真解，即要求当$h\to0$，$\tau\to0$时，有

$$|\tilde{u}_j^k-u(x_j,t_k)|\leqslant|\tilde{u}_j^k-u_j^k|+|u_j^k-u(x_j,t_k)|\to0.$$

要$u_j^k \to u(x_j, t_k)$，这是**收敛性问题**；要$\tilde{u}_j^k \to u_j^k$，是差分方程的**稳定性问题**，即在初边值有干扰的情况下，误差$|\tilde{u}_j^k - u_j^k|$能被有效地控制。

由于齐次方程的初值稳定性具有代表性（参见§1），下面以（1.18）式相应的齐次初值问题的古典显格式

$$u_j^{k+1} = u_j^k + r(u_{j-1}^k - 2u_j^k + u_{j+1}^k)$$

的数值计算为例，观察误差$e_j^k = \tilde{u}_j^k - u_j^k$的传播情形。由上式显见，误差$e_j^k$满足误差方程

$$e_j^{k+1} = e_j^k + r(e_{j-1}^k - 2e_j^k + e_{j+1}^k).$$

假设$k-1$层之前的计算没有误差，即$e_j^{k-1} = 0$；而在第k层仅于x_{j_0}处有计算误差$e_{j_0}^k = \varepsilon$，这一层其它点也无误差，且在计算过程中不再产生新的误差；那么利用误差方程，则可计算出$r = 1/2$和$r = 1$时的误差ε的传播情况，并将其列于表1.1和表1.2中。

表1.1　$r = 1/2$时，古典显格式的误差传播

k \ j	j_0-4	j_0-3	j_0-2	j_0-1	j_0	j_0+1	j_0+2	j_0+3	j_0+4
k					ε				
$k+1$				0.5ε	0	0.5ε			
$k+2$			0.25ε	0	0.5ε	0	0.25ε		
$k+3$		0.125ε	0	0.375ε	0	0.375ε	0	0.125ε	
$k+4$	0.0625ε	0	0.25ε	0	0.375ε	0	0.25ε	0	0.0625ε

由表1.1中的数值可见，随着k增大，误差ε的影响逐渐变小，所以古典显格式在$r = 1/2$时是关于初值稳定的。然而，由表1.2可见，随着k增大，误差ε的影响越来越大，古典显格式在$r = 1$时是关于初值不稳定的，此时的差分解也是不收敛的。那么，如何分析判断某种差分格式的稳定性？稳定性与收敛性的关

系如何?

由此可见，差分格式的稳定性和差分解的收敛性，是解抛物型方程的差分方法中必须首先弄清楚的两个基本问题。

表1.2 $r=1$ 时，古典显格式的误差传播

k \ j	j_0-4	j_0-3	j_0-2	j_0-1	j_0	j_0+1	j_0+2	j_0+3	j_0+4
k					ε				
k+1				ε	$-\varepsilon$	ε			
k+2			ε	0	ε	0	ε		
k+3		ε	$-\varepsilon$	2ε	$-\varepsilon$	2ε	$-\varepsilon$	ε	
k+4	ε	-2ε	2ε	-4ε	3ε	-4ε	2ε	-2ε	ε

§3 差分方程的稳定性与收敛性

本节我们将要明确有关稳定性和收敛性的概念并建立有关理论。

3.1 差分问题是对连续性问题的近似模拟

设在t_k时间层差分方程的解是 $U^k=(u_1^k,\cdots,u_{N-1}^k)^T$，它可看作是定义在$t_k$时间层网格节点上的解向量，也可看作是**网格函数** $u_h^k(x)$ (例如阶梯函数：$u_h^k(x)=u_j^k$，$x_{j-1/2}<x\leqslant x_{j+1/2}$)在网格节点上的表现。显然，差分解$U^k$或$u_h^k(x)$是对抛物型方程的连续解$u^k(x)$的近似模拟。

对于差分解向量 U^k，我们经常使用**欧氏 (Euclid) 模**和**最大模**：

$$\|U^k\|_2^2=h\sum_j|u_j^k|^2,\quad \|U^k\|_\infty=\max_j|u_j^k|.$$

显然，视差分解向量U^k为阶梯函数的表现时，$\|U^k\|_0$确为$u_h^k(x)$的L_2模；当$h\to 0$时，它趋近于方程解$u^k(x)$的L_2模。因此，差分解的模$\|U^k\|_0$和$\|U^k\|_\infty$是对解函数$u^k(x)$的L_2模和连续模的一种近似模拟。

易知，差分解U^k按上述某种模$\|\cdot\|$收敛于方程解$u^k=(u(x_1,t_k),\cdots,u(x_{N-1},t_k))^T$，可描述为

$$\|U^k-u^k\|\to 0,\qquad \tau,h\to 0,\ k\tau\to t.$$

由§2知，截断误差$R_j^k=L_h[u]_j^k-[Lu]_j^k$反映了差分算子和微分算子的关系，据此有

相容逼近定义 如果当$\tau,h\to 0,k\tau\to t$时，对于充分光滑函数u，有

$$\|R_j^k\|=\|L_h[u]_j^k-[Lu]_j^k\|\to 0,$$

则称L_h是对L的相容逼近，或简称差分方程与微分方程**相容**。

显然，古典显格式(1.20)、古典隐格式(1.23)和Richardson格式(1.26)都是对热传导方程(1.18)的相容逼近。

下面观察差分方程解的结构。由古典显格式(1.22)和古典隐格式(1.25)不难知道，常系数线性微分方程的两层差分格式可表示为

$$A(h,\tau)U^{k+1}=B(h,\tau)U^k+\tau F^k \tag{1.29}$$

的形式，其中A、B是常数矩阵，它们的元素只与h和τ有关，与x和t无关，这是由于微分方程是线性常系数和差分方法中用差商直接代替微商的缘故。

对于变系数线性方程，相应的两层格式可写成

$$A^{(k)}U^{k+1}=B^{(k)}U^k+\tau F^k,$$

其中$A^{(k)}$和$B^{(k)}$的元素不仅与h、τ有关，而且与时间层k有关。

值得指出的是，差分方程中h和τ总是有一定联系的。一方面，描述差分解$u_h^k(x)$逼近方程解$u^k(x)$时，要求$h\to 0$且$\tau\to 0$；另一方面，差分方程的稳定性与"**网比**"$r=a^2\frac{h^2}{\tau}$有关(参见表

1.1 和表 1.2 以及以后关于稳定性分析的章节)；所以差分方程中 h 和 τ 的关系可用连续函数 $h=g_r(\tau)$ 描述之，且要求 $g_r(0)=0$。

若设(1.29)式中的$A^{-1}(h,\tau)$存在，且记 $\sigma(h,\tau)=A^{-1}(h,\tau)\cdot B(h,\tau)$，则常系数两层差分格式(1.29)的解具有如下结构

$$\begin{aligned}U^{k+1}&=\sigma(h,\tau)U^k+\tau A^{-1}(h,\tau)F^k\\&=\sigma(h,\tau)[\sigma(h,\tau)U^{k-1}+\tau A^{-1}(h,\tau)F^{k-1}]+\tau A^{-1}(h,\tau)F^k\cdots\\&=\sigma^{k+1}(h,\tau)U^0+\tau\sum_{l=0}^{k}\sigma^{k-l}(h,\tau)\cdot A^{-1}(h,\tau)\cdot F^l.\end{aligned}\quad(1.30)$$

由此可见，二阶常系数线性抛物型方程的两层差分格式的解结构类似于定解问题的解结构(1.9)或(1.15)式.

3.2 差分方程的稳定性

仿照抛物型方程的理论研究(参见§1)，利用差分方程的解结构(1.30)式来研究差分方程的稳定性是非常方便的.

差分方程的初值稳定性是研究差分方程在右端自由项无误差(相当于齐次差分方程)的情况下，初值干扰对差分方程解的影响. 它反映了差分解是否连续依赖于初值的情形. 只要在(1.30)式中令$F^k\equiv0$，即可理解下述关于初值稳定的定义.

定义 称差分格式 (1.29) 是关于$\|\cdot\|$**初值稳定**的，若存在$M>0$，$\tau_0>0$，使得当 $0<\tau<\tau_0$，$h=g_r(\tau)$ 且 $g_r(0)=0$ 时，与(1.29)式相应的齐次差分方程的解满足不等式

$$\|U^k\|\leqslant M\|U^0\|,\quad 0<\forall k<T/\tau,$$

其中，M是与 τ 无关的常数.

差分方程的右端稳定性是研究差分方程在初值准确无误差的情况下，右端自由项 F^k 的干扰对差分解的影响，它反映了差分解是否连续依赖于右端自由项的情形，这相当于研究非齐次差分方程与零初值条件的情形. 因此，只要在(1.30)式中令 $U^0\equiv0$，即可理解右端稳定的定义.

定义 差分方程(1.29)称为关于模$\|\cdot\|$**右端稳定**的，若存在$M>0$，$\tau_0>0$，使得当$0<\tau<\tau_0$，$h=g_r(\tau)$且$g_r(0)=0$时，差分方程(1.29)适合初值条件$U^0=0$的解满足不等式

$$\|U^k\|\leqslant M\tau\sum_{l=0}^{k-1}\|F^l\|,\qquad 0<\forall k<T/\tau,$$

其中，M是与τ无关的常数.

定理1.1 差分方程按$\|\cdot\|$初值稳定的充要条件是，存在$\tau_0>0$，$M>0$，使得当$0<\tau<\tau_0$，$h=g_r(\tau)$且$g_r(0)=0$时，有

$$\|\sigma^k(h,\tau)\|\leqslant M,\qquad 0<\forall k<T/\tau, \tag{1.31}$$

式中，$\sigma(h,\tau)=A^{-1}(h,\tau)B(h,\tau)$，$A^{-1}$存在.

证明 令(1.30)式中$F^l\equiv 0$，并两边取模，有

$$\|U^k\|=\|\sigma^k(h,\tau)U^0\|\leqslant\|\sigma^k(h,\tau)\|\|U^0\|,$$

再由初值稳定的定义，即证得定理1.1. □

定理1.2 若差分方程(1.29)关于$\|\cdot\|$初值稳定，则其关于$\|\cdot\|$右端稳定.

证明 差分方程(1.29)的初值稳定性和右端稳定性，可分别研究差分方程

$$\begin{cases}AU_1^{k+1}=BU_1^k;\\U_1^0=U_1^0,\end{cases}\qquad 和\qquad \begin{cases}AU^{k+1}=BU^k+\tau F^k;\\U^0=0,\end{cases}$$

由(1.30)式知，这两个方程的解分别是

$$U_1^{k+1}=\sigma^{k+1}(h,\tau)U_1^0;$$

$$U^{k+1}=\tau\sum_{l=0}^{k}\sigma^{k-l}(h,\tau)A^{-1}(h,\tau)F^l.$$

因差分方程(1.29)关于$\|\cdot\|$初值稳定，利用定理1.1即有$\|A^{-1}(h,\tau)\|<C$，$\|\sigma^k(h,\tau)\|<M\,(0<\forall k<T/\tau)$，故证得

$$\|U^{k+1}\|\leqslant\tau CM\sum_{l=0}^{k}\|F^l\|.\qquad □$$

综上所述，类似于抛物型方程稳定性的讨论，差分方程稳定性的讨论可归结为初值稳定性讨论. 因此，在以后章节中，差分

方程的稳定性都是指相应齐次方程的初值稳定性；同样为了方便起见，一律将差分格式写成齐次形式。

3.3 收敛性定理

定理1.3 设线性常系数抛物型方程$Lu=f$与两层差分格式(1.29)相容，且(1.29)式关于$\|\cdot\|$稳定，则差分解按$\|\cdot\|$收敛于方程的解，即

$$\|U^k-[u]^k\|\to 0,\quad h、\tau\to 0,\ k\tau\to t.$$

如果截断误差阶为$R_j^k=O(\tau^\alpha+h^\beta)$，则有误差估计

$$\|U^k-[u]^k\|=O(\tau^\alpha+h^\beta).$$

证明 考虑初值问题，令

$$e_j^k=u(x_j,t_k)-u_j^k=[u]_j^k-u_j^k,\quad e_j^0=0,$$

$$E^k=(e_1^k,\cdots,e_{N-1}^k)^T,\quad R^k=(R_1^k,\cdots,R_{N-1}^k)^T.$$

由于

$$R_j^k=L_h[u]_j^k-[Lu]_j^k=L_h[u]_j^k-f_j^k$$
$$=L_h[u]_j^k-L_hu_j^k=L_he_j^k,$$

故误差e_j^k满足差分方程

$$\begin{cases}L_he_j^k=R_j^k,\\ e_j^0=0,\end{cases}$$

其两层差分格式形如(1.29)式，且由(1.30)式知其差分解为

$$E^k=\tau\sum_{l=0}^{k-1}[\sigma(h,\tau)]^{k-l}\cdot A^{-1}(h,\tau)R^l.$$

由初值稳定性导出

$$\|A^{-1}(h,\tau)\|\leqslant M,\quad \|\sigma^i(h,\tau)\|\leqslant M,\qquad i=1,\cdots,k,$$

故有

$$\|E^k\|\leqslant \tau M^2\sum_{l=0}^{k-1}\|R^l\|=\tau\cdot O(\tau^\alpha+h^\beta)=O(\tau^\alpha+h^\beta).\qquad \square$$

值得注意的是，定理1.3中对模无特殊要求，所以，只要证明差分格式按确定模稳定，就可导出差分解按相应模收敛。另

外，定理 1.3 中的误差估计不仅说明了差分解的收敛性，而且指出了提高差分解精度的关键之一在于提高截断误差的阶。定理1.3还可进一步发展为较一般情形。

Lax 等价定理 设线性微分方程关于初值是适定的，其相应的差分方程满足相容逼近条件，则差分方程收敛的充要条件是差分方程关于初值稳定。此定理的证明从略，有兴趣的读者可参见参考文献[1]～[4]。

最后，我们把本节讨论的各种概念之间的关系简单表示在图1.5中。

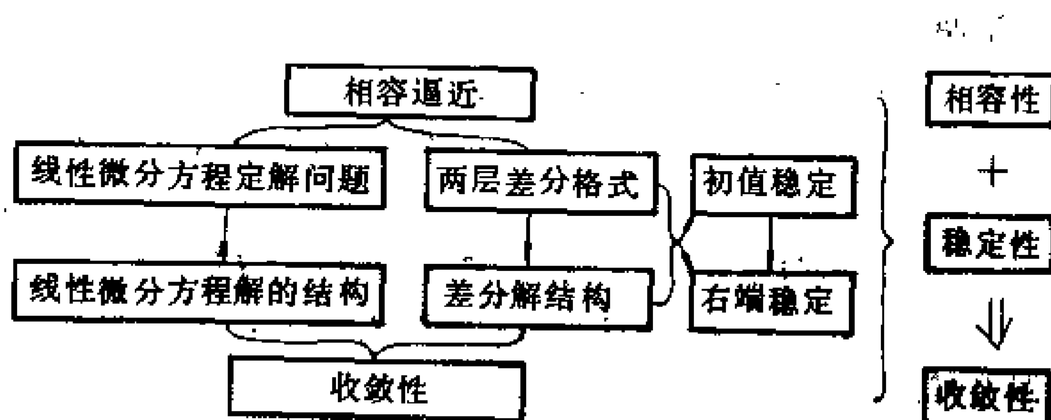


图 1.5

§4 判别稳定性的Fourier方法

首先考虑线性常系数抛物型方程初边值问题的两层差分格式的稳定性。为便于分析，我们把两层差分格式写成形如(1.20)和(1.23)式的形式。一般地，两层差分格式可统一写成

$$\sum_{j\in\Omega_1} a_j u_{m+j}^{k+1} = \sum_{j\in\Omega_0} b_j u_{m+j}^{k} + \tau f_m^k$$

的形式，它用于稳定性分析的齐次差分方程是

$$\sum_{j\in\Omega_1} a_j u_{m+j}^{k+1} = \sum_{j\in\Omega_0} b_j u_{m+j}^{k}, \tag{1.32}$$

式中，Ω_1和Ω_0分别表示在第$k+1$层和第k层上那些j可取值的集合。例如，对古典显格式(1.20)，$\Omega_1=\{0\}, \Omega_0=\{-1,0,1\}$，$a_0=1$，$b_{-1}=b_1=r$，$b_0=1-2r$。

对于给定的网格节点值$u_j^k (j=0,\cdots,N)$，固定k，可以定义函数$u_h^k(x)$，$x\in[0,1]$。例如，$u_h^k(x)$可以是阶梯函数，它在$x_j-h/2<x\leqslant x_j+h/2 (j=1,\cdots,N-1)$上取值$u_j^k$，在$[0,h/2]$和$[1-h/2,1]$上取值$u_0^k=u_N^k=0$(假定齐次边值是不失一般性的)。所以，我们可以把$u_h^k(x)$周期地开拓到整个实数轴上。

利用正交函数系$\{e^{i\pi l x}\}$，将$u_h^k(x)$展成Fourier级数：

$$u_h^k(x)=\sum_{l=-\infty}^{+\infty} v^k(l) e^{i\pi l x},$$

其中$v^k(l)$是系数，$v^k(l)=\dfrac{1}{2}\int_{-1}^{1} u_h^k(\xi) e^{-i\pi l \xi} d\xi$。根据Parseval等式

$$\|u_h^k(x)\|_0^2=\int_0^1 [u_h^k(x)]^2 dx=\sum_{l=-\infty}^{+\infty} |v^k(l)|^2 dx,$$

知$u_h^k(x)$的L_2模与$U^k=(u_1^k,u_2^k,\cdots,u_{N-1}^k)^T$的Euclid模是等价的。

利用开拓后的$u_h^k(x)$，(1.32)式可以写成

$$\sum_{j\in\Omega_1} a_j u_h^{k+1}(x+jh)=\sum_{j\in\Omega_0} b_j u_h^k(x+jh),$$

从而有

$$\sum_{j\in\Omega_1} a_j\left[\sum_{l=-\infty}^{+\infty} v^{k+1}(l) e^{i\pi l(x+jh)}\right]$$
$$=\sum_{j\in\Omega_0} b_j\left[\sum_{l=-\infty}^{+\infty} v^k(l) e^{i\pi l(x+jh)}\right],$$

整理后有

$$\sum_{l=-\infty}^{+\infty} e^{i\pi l x}\left[\sum_{j\in\Omega_1} a_j e^{i\pi j l h}\right] v^{k+1}(l)$$
$$=\sum_{l=-\infty}^{+\infty} e^{i\pi l x}\left[\sum_{j\in\Omega_0} b_j e^{i\pi j l h}\right] v^k(l).$$

注意$\{e^{i\pi lx}\}$的正交性，即

$$\int_{-1}^{1} e^{i\pi lx} e^{i\pi mx} dx = \begin{cases} 0, & \text{当 } l \neq m \text{时,} \\ 1, & \text{当 } l = m \text{时,} \end{cases}$$

则由前式导出

$$\left[\sum_{j \in \Omega_1} a_j e^{i\pi jlh}\right] v^{k+1}(l) = \left[\sum_{j \in \Omega_0} b_j e^{i\pi jlh}\right] v^k(l),$$

$$l = 0, \ \pm 1, \ \pm 2, \ \cdots$$

若记 $\sigma = \pi l$，且记（因$h = g_r(\tau)$）

$$G(\sigma, \tau) = \left[\sum_{j \in \Omega_1} a_j e^{i\sigma jh}\right]^{-1} \left[\sum_{j \in \Omega_0} b_j e^{i\sigma jh}\right],$$

则上式可改写为

$$v^{k+1}(l) = G(\sigma, \tau) v^k(l) \quad (l = 0, \pm 1, \cdots). \tag{1.33}$$

$G(\sigma,\tau)$称为传播因子。因为$G(\sigma,\tau)$只与σ、h和τ有关，而与时间层次k无关，则反复利用上式可导出

$$v^k(l) = G^k(\sigma, \tau) v^0(l). \tag{1.34}$$

由此可见，差分方程（1.32）稳定性的讨论就转化为方程(1.34)稳定性的讨论。

定理1.4 差分方程(1.32)按L_2模稳定的充要条件是：存在M、$\tau_0 > 0$，对$0 < \tau \leqslant \tau_0$、$0 < \forall k < T/\tau$和$\forall \sigma$有

1° $|G^k(\sigma, \tau)| \leqslant M$,

或

2° $|G(\sigma, \tau)| \leqslant 1 + c\tau$ （常数c与h、τ无关）。 (1.35)

(1.35)式亦称为von Neumann条件。

证明 先就条件1°证明之。

充分性 设条件1°成立，即有

$$\sum_{l=-\infty}^{+\infty} |v^k(l)|^2 = \sum_{l=-\infty}^{+\infty} |G^k(\sigma,\tau) v^0(l)|^2 \leqslant M^2 \sum_{l=-\infty}^{+\infty} |v^0(l)|^2,$$

再利用 Parseval 等式，即证得

$$\|u_h^k(x)\|_0^2 \leqslant M^2 \|u_h^0(x)\|_0^2.$$

必要性 设差分格式关于L_2模稳定，即上式成立，则有

$$\sum_{l=-\infty}^{+\infty} |G^k(\sigma,\tau)|^2 \cdot |v^0(l)|^2 \leqslant M^2 \sum_{l=-\infty}^{+\infty} |v^0(l)|^2,$$

再利用反证法(略)即可证得1°。

最后证明条件1°和2°的等价性。事实上，由 1° 知 $|G(\sigma,\tau)| \leqslant M^{1/k}$对$0<\forall k<T/\tau$成立，特别取$(T-\tau)/\tau\leqslant k<T/\tau$，则有

$$|G(\sigma,\tau)| \leqslant M^{\frac{\tau}{T-\tau}} = \exp\left[\frac{\tau}{T-\tau}\ln M\right]$$

$$\leqslant 1+\frac{\tau}{T-\tau}\ln M+\frac{1}{2!}\left(\frac{\tau}{T-\tau}\ln M\right)^2+\cdots$$

$$\leqslant 1+\frac{\tau}{T-\tau}\ln M\exp\left[\frac{\tau}{T-\tau}\ln M\right]\leqslant 1+c\tau,$$

其中 $c=\dfrac{\ln M}{T-\tau_0}\exp\left[\dfrac{\tau_0}{T-\tau_0}\ln M\right]$. □

判定稳定性的 Fourier 分析法要求在任一时间层 $k\geqslant 0$ 上，解函数 $u(x,t_k)$ 可以关于变量 x 周期延拓到整个实轴上并写成 Fourier 级数的形式，这是同数理方程中的分离变量法一致的。该方法也适用于研究 Cauchy问题差分格式的稳定性，对于这种情形，$u(x,t_k)$定义在整个实数轴上，但一般说来不具有周期性，此时只要将 $u(x,t_k)$ 展开成 Fourier 积分形式，仍可导出一个方程式情形差分格式稳定的充要条件是von Neumann条件。

判定稳定性的Fourier 分析法使用起来也是很方便的，由该方法的推导过程可知，实际使用该方法时，只要将 u_j^k 改写为 $u_h^k(x)$，并取$u_h^k(x)=v^k(l)e^{i\sigma x}$，再将再代入差分方程，消去公因子即得到(1.33)式，从而求得$G(\sigma,\tau)$。

还需注意，在使用von Neumann条件判别差分格式稳定性时，由于(1.35)式中的c 难以确定，通常利用 $|G(\sigma,\tau)|\leqslant 1$ 作为判别稳定性的充分条件。

作为例子，判别古典显格式(1.20)的稳定性。

先将古典显格式写为

$$u_h^{k+1}(x)=(1-2r)u_h^k(x)+r(u_h^k(x-h)+u_h^k(x+h)),$$

再将$u_h^k(x)=v^k(l)e^{i\sigma x}$代入上式得

$$e^{i\sigma x}v^{k+1}(l)=[(1-2r)+re^{-i\sigma h}+re^{i\sigma h}]e^{i\sigma x}v^k(l),$$

消去公因子$e^{i\sigma x}$，有

$$v^{k+1}(l)=[(1-2r)+r(e^{-i\sigma h}+e^{i\sigma h})]v^k(l),$$

从而得传播因子

$$G(\sigma,\tau)=1-2r(1-\cos\sigma h)=1-4r\sin^2\frac{\sigma h}{2}.$$

因此，若令$r=a^2\tau/h^2=$常数，则由上式推出，当$r\leqslant 1/2$时满足von Neumann条件，即当$r\leqslant 1/2$时古典显格式才稳定。由此可以解释，$r=1$时（见表1.2），古典显格式不稳定的原因。

判别古典隐格式(1.23)的稳定性。

将古典隐格式改写为

$$(1+2r)u_h^{k+1}(x)-ru_h^{k+1}(x-h)-ru_h^{k+1}(x+h)=u_h^k(x),$$

令$u_h^k(x)=v^k(l)e^{i\sigma x}$，且将其代入上式得

$$[1+2r-r(e^{-i\sigma h}+e^{i\sigma h})]e^{i\sigma x}v^{k+1}(l)=e^{i\sigma x}v^k(x),$$

$$[1+2r-r(e^{-i\sigma h}+e^{i\sigma h})]v^{k+1}(l)=v^k(x),$$

从而得传播因子$G(\sigma,\tau)=1\Big/\left(1+4r\sin^2\dfrac{\sigma h}{2}\right)$.

因此，对一切网比r，都有$|G(\sigma,\tau)|\leqslant 1$，满足von Neumann条件，古典隐格式(1.23)**绝对稳定（无条件稳定）**。

Fourier 分析方法也适用于判断解微分方程组的差分格式（差分方程组）的稳定性。对于微分方程组

$$\frac{\partial U}{\partial t}=C\frac{\partial^2 U}{\partial x^2},$$

其中，U是$n\times 1$阶矩阵，C是$n\times n$阶常系数矩阵；其相应的两层差分格式也可统一写成

$$\sum_{l\in\Omega_1} A_l U_{j+l}^{k+1} = \sum_{l\in\Omega_0} B_l U_{j+l}^{k} \tag{1.36}$$

的形式，其中，A_l、B_l都是常系数矩阵，U_j^k是$U(x_j,t_k)$的近似值。我们同样可以利用U_j^k定义$U_h^k(x)$，并开拓到整个数轴上去，再作傅氏展开，最后就得到形如(1.34)式子：

$$V^{k+1}(l)=G^{k+1}(\sigma,\tau)V^0(l).$$

不过，这里的$G(\sigma,\tau)$是矩阵，称为**传播矩阵**。

定理1.5 1° 两层差分格式(1.36)稳定的充要条件是：

$\exists M$、$\tau_0>0$，对$0<\tau<\tau_0$、$0<\forall k<T/\tau$和$\forall\sigma$有

$\|G^k(\sigma,\tau)\|\leqslant M$ （常数M与h和τ无关）。

2° 两层差分格式稳定的必要条件是：

$$\rho(G)\leqslant 1+c\tau,$$

其中$\rho(\cdot)$表示**谱半径**，常数c与h和τ无关，上式也称为Von Neumann条件。

3° 若G是**正规矩阵**，则Von Neumann条件也是差分格式(1.36)按L_2模稳定的充分条件。

证明 1°的证明是显然的，证明从略。

2°的证明：因为谱半径是最小的一种模量，即$\rho(G)\leqslant\|G\|$，所以利用1°就可证得2°。

3°的证明主要利用$\rho(G)=\|G\|_0=\|G\cdot G^*\|_0^{1/2}$，这个等式对$G$是正规阵成立，故由1°知3°成立。 □

值得注意的是，对于1×1阶矩阵G，因$\rho(G)=|G|$，所以本定理中的Von Neumann条件是与(1.35)式一致的，它是差分格式稳定的充要条件；但对于G是矩阵情形，Von Neumann条件只是差分格式稳定的必要条件。另外，本定理中的3°是验证差分格式稳定的依据，虽然它限制G是正规阵，但其应用范围还是很

广泛的，因为正规阵G的定义是$GG^*=G^*G$，故知**实对称阵**、**Hermite矩阵**、**正交矩阵**和**U矩阵**都是正规矩阵。

判别差分格式稳定，还有如下定理。

定理1.6 若存在相似变换$S(\sigma,\tau)$，且对一切σ和$0<\tau<\tau_0$，有

$$S(\sigma,\tau)G(\sigma,\tau)S^{-1}(\sigma,\tau)=D(\sigma,\tau),$$

$$\|S(\sigma,\tau)\|\leqslant c_1,\quad \|S^{-1}(\sigma,\tau)\|\leqslant c_2,$$

其中，$D(\sigma,\tau)$是与$G(\sigma,\tau)$同阶的对角阵，则 Von Neumann 条件是差分格式(1.36)稳定的充分条件。

证明 由定理条件有$\rho(G)=\rho(D)$ 且

$$G(\sigma,\tau)=S(\sigma,\tau)D(\sigma,\tau)S^{-1}(\sigma,\tau),$$

连续使用上式，并取模，就有

$$\begin{aligned}\|G^k(\sigma,\tau)\|&=\|S(\sigma,\tau)D^k(\sigma,\tau)S^{-1}(\sigma,\tau)\|\\&\leqslant M\rho^k(D(\sigma,\tau)).\end{aligned}$$

再利用Von Neumann条件$\rho(G)=\rho(D)\leqslant 1+O(\tau)$，即知$\|G^k\|$一致有界，由定理1.5中的1°，即证得差分格式稳定。 □

定理1.7 设传播矩阵$G(\sigma,\tau)=\hat{G}(w)$，其中$w=\sigma h$，$\hat{G}(w)$是$n\times n$阶矩阵。

1° 若$\hat{G}(w)$有n个不同的特征值，则 Von Neumann 条件是差分格式(1.36)稳定的充分条件

2° 若$\rho(\hat{G}(w))<1$，则差分格式(1.36)稳定。

证明 1° 因$\hat{G}(w)$有n个不同的特征值，所以存在非奇异矩阵$S(w)$，使有

$$S(w)\hat{G}(w)S^{-1}(w)=D(w),$$

其中，$D(w)$是对角阵。由定理1.6即可证得 Von Neumann 条件是差分格式(1.36)稳定的充分条件。

2° 据线性代数知识，若设A是$n\times n$阶矩阵且$\rho(A)<1$，那

么必存在某种矩阵模$\|\cdot\|_S$，使$\|A\|_S<1$和$\|B\cdot C\|_S\leqslant\|B\|_S\|C\|_S$（$B$、$C$是与$A$同阶的矩阵）。对于任一个$w_0\in R$，因$\rho(\hat{G}(w_0))<1$，那么就有$\|\hat{G}(w_0)\|_S=L<1$。人们还可选择足够小的$\varepsilon>0$，使得对$\forall w\in(w_0-\varepsilon,\ w_0+\varepsilon)$，有

$$\|\hat{G}(w)\|_S\leqslant\frac{1}{2}(L+1)<1,$$

$$\|\hat{G}^k(w)\|_S\leqslant\frac{1}{2}(L+1)<1\quad(k\in N).$$

最后利用矩阵模的等价性，就有

$$\|\hat{G}^k(w)\|_0\leqslant M\|\hat{G}^k(w)\|_S<M$$
$$(k\in N,\ w\in(w_0-\varepsilon,\ w_0+\varepsilon)).\quad\square$$

作为差分方程组稳定性分析的例子，我们判别 **Richardson** 格式（1.26）的稳定性。因为（1.26）式是一个三层格式，故不能直接应用 **Fourier** 分析方法。为此，先将其化为差分方程组

$$\begin{pmatrix}u_j^{k+1}\\ u_j^k\end{pmatrix}=\begin{pmatrix}2r&0\\0&0\end{pmatrix}\begin{pmatrix}u_{j-1}^k\\ u_{j-1}^{k-1}\end{pmatrix}+\begin{pmatrix}-4r&1\\1&0\end{pmatrix}\begin{pmatrix}u_j^k\\ u_j^{k-1}\end{pmatrix}$$
$$+\begin{pmatrix}2r&0\\0&0\end{pmatrix}\begin{pmatrix}u_{j+1}^k\\ u_{j+1}^{k-1}\end{pmatrix},$$

令$W_j^{k+1}=(u_j^{k+1},u_j^k)^T$且扩展为$W^{k+1}(x)$，则上式可改写成为

$$W^{k+1}(x)=\begin{pmatrix}2r&0\\0&0\end{pmatrix}W^k(x-h)$$
$$+\begin{pmatrix}-4r&1\\1&0\end{pmatrix}W^k(x)+\begin{pmatrix}2r&0\\0&0\end{pmatrix}W(x+h),$$

再令$W^k(x)=V^k(l)e^{i\sigma x}$，代入上式并消去公因子$e^{i\sigma x}$，经整理得

$$V^{k+1}(l)=G(\sigma,\tau)V^k(l),$$

其中，

$$G(\sigma,\tau)=\begin{pmatrix} -8r\sin^2\frac{\sigma h}{2} & 1 \\ 1 & 0 \end{pmatrix}.$$

求出传播矩阵G的两个特征值

$$\lambda_{1,2}=-4r\sin^2\frac{\sigma h}{2}\pm\sqrt{16r^2\sin^4\frac{\sigma h}{2}+1};$$

但当 $\sin\frac{\sigma h}{2}$充分小时，

$$\rho(G)=\max|\lambda_{1,2}|=4r\sin^2\frac{\sigma h}{2}+\left(16r^2\sin^4\frac{\sigma h}{2}+1\right)^{1/2}$$

$$>4r\sin^2\frac{\sigma h}{2}+1+\left(16r^2\sin^4\frac{\sigma h}{2}\right)^{1/2}>1;$$

即对任何网比 r，$\rho(G)>1$，不满足 Von Neumann 条件，故 Richardson 格式 **绝对不稳定**。

§5 其它差分格式及其稳定性分析

由§3和§4的结论可知：提高截断误差阶，设计高精度的差分格式；设计稳定性好的差分格式，是解抛物型方程差分方法的两个核心问题。本节将重点就提高截断误差阶和改善稳定性两个方面，介绍几种求解(1.18)式的实用的差分格式，并用 Fourier分析法讨论其稳定性。

5.1 加权隐格式和Crank-Nicholson格式

求解(1.18)式的古典显格式是条件稳定的，古典隐格式是绝对稳定的，它们的截断误差都是 $O(\tau+h^2)$。这两种格式的加权平均格式是否能在稳定性和截断误差方面有所改善呢？

为此，用 $\theta(0\leqslant\theta\leqslant1)$ 乘古典隐格式，用 $(1-\theta)$ 乘古典显格式，相加得

$$u_j^{k+1}-u_j^k=r[\theta(u_{j-1}^{k+1}-2u_j^{k+1}+u_{j+1}^{k+1})$$
$$+(1-\theta)(u_{j-1}^k-2u_j^k+u_{j+1}^k)]. \qquad (1.37)$$

这个格式称为**加权隐格式**，其节点图式见图1.6.

截断误差分析。因为

$$\frac{1}{\tau}[u_j^{k+1}-u_j^k]=\left[\frac{\partial u}{\partial t}\right]_j^{k+1/2}+O(\tau^2),$$

$$\frac{1}{h^2}[u_{j-1}^{k+1}-2u_j^{k+1}+u_{j+1}^{k+1}]=\left[\frac{\partial^2 u}{\partial x^2}\right]_j^{k+1}+O(h^2)$$

$$=\left[\frac{\partial^2 u}{\partial x^2}\right]_j^{k+1/2}+\frac{\tau}{2}\left[\frac{\partial^3 u}{\partial x^2\partial t}\right]_j^{k+1/2}+O(\tau^2+h^2),$$

$$\frac{1}{h^2}[u_{j-1}^k-2u_j^k+u_{j+1}^k]$$

$$=\left[\frac{\partial^2 u}{\partial x^2}\right]_j^{k+1/2}-\frac{\tau}{2}\left[\frac{\partial^3 u}{\partial x^2\partial t}\right]_j^{k+1/2}+O(\tau^2+h^2),$$

故知，加权隐格式的截断误差为

$$R_j^k=a^2\tau\left(\frac{1}{2}-\theta\right)\left[\frac{\partial^3 u}{\partial x^2\partial t}\right]_j^{k+1/2}+O(\tau^2+h^2). \qquad (1.38)$$

稳定性分析。用 Fourier 分析方法很容易求得传播因子

$$G(\sigma,\tau)=\left[1-4(1-\theta)r\sin^2\frac{\sigma h}{2}\right]\Big/\left[1+4\theta r\sin^2\frac{\sigma h}{2}\right].$$

据定理1.4知，von Neumann条件是加权隐格式(1.37)按L_2模稳定的充要条件。$-1\leqslant G(\sigma,\tau)\leqslant 1$ 的右边不等式恒成立。现考虑左边不等式

$$-\left(1+4\theta r\sin^2\frac{\sigma h}{2}\right)\leqslant 1-4(1-\theta)r\sin^2\frac{\sigma h}{2},$$

即是

$$4r(1-2\theta)\sin^2\frac{\sigma h}{2}\leqslant 2,$$

所以，加权隐格式(1.37)稳定性条件是

$$\begin{cases} r\leqslant 1/2(1-2\theta), & \text{当}\, 0\leqslant\theta<1/2\text{时} \\ \text{绝对稳定}, & \text{当}\, \theta\geqslant 1/2\text{时}. \end{cases} \tag{1.39}$$

总之，加权隐格式包括了古典显格式($\theta=0$)和古典隐格式 ($\theta=1$) 的情形。特别地，当$\theta=1/2$时，由(1.38)式和(1.39)式，有截断误差为 $O(\tau^2+h^2)$ 的差分格式

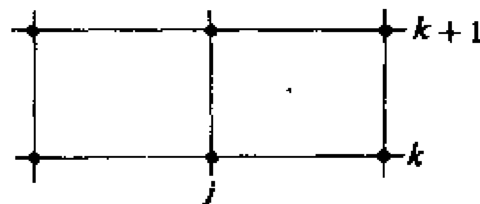


图 1.6

$$u_j^{k+1}-u_j^k=\frac{1}{2}r[(u_{j-1}^{k+1}-2u_j^{k+1}+u_{j+1}^{k+1}) + (u_{j-1}^k-2u_j^k+u_{j+1}^k)]. \tag{1.40}$$

此格式称作 **Crank-Nicholson 格式**(1947年)，这是一个实用的绝对稳定格式。

5.2 Du Fort-Frankel 格式

用$\frac{1}{2}(u_j^{k+1}+u_j^{k-1})$代替 Richardson 格式(1.26)中的$u_j^k$，有

$$\frac{u_j^{k+1}-u_j^{k-1}}{2\tau}-a^2\frac{u_{j-1}^k-(u_j^{k+1}+u_j^{k-1})+u_{j+1}^k}{h^2}=0, \tag{1.41}$$

此格式称作 **Du Fort-Frankel 格式** (1954年)。这是一个三层显格式，其节点关系见图1.7.

先考察此格式的相容性。因

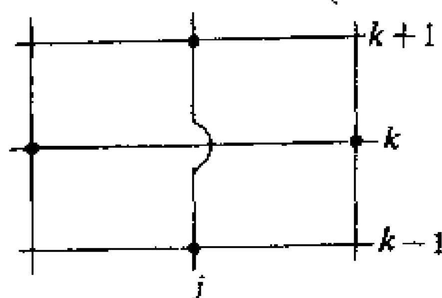


图 1.7

$$\frac{u_j^{k+1}-u_j^{k-1}}{2\tau}=\left[\frac{\partial u}{\partial t}\right]_j^k+O(\tau^2),$$

$$a^2\frac{u_{j-1}^k-2u_j^k+u_{j+1}^k}{h^2}=a^2\left[\frac{\partial^2 u}{\partial x^2}\right]_j^k+O(h^2),$$

$$-a^2\frac{u_j^{k-1}-2u_j^k+u_j^{k+1}}{\tau^2}\left(\frac{\tau}{h}\right)^2=-a^2\left(\frac{\tau}{h}\right)^2\left\{\left[\frac{\partial^2 u}{\partial t^2}\right]_j^k+O(\tau^2)\right\},$$

故有

$$R_j^k=a^2\left(\frac{\tau}{h}\right)^2\left(\frac{\partial^2 u}{\partial t^2}\right)_j^k+O(\tau^2+h^2)+O\left(\frac{\tau^4}{h^2}\right).$$

由此可见，当$\tau/h\to 0$，即τ比h趋近于零要快些时，**Du Fort-Frankel** 格式相容于$\frac{\partial u}{\partial t}=a^2\frac{\partial^2 u}{\partial x^2}+f$且截断误差为$O(\tau^2+h^2)$；当$\tau/h$保持常数$\beta$，则 **Du Fort-Frankel** 格式相容于双曲型方程$\frac{\partial u}{\partial t}-a^2\frac{\partial^2 u}{\partial x^2}+a^2\beta^2\frac{\partial^2 u}{\partial t^2}=f$，截断误差也是$O(\tau^2+h^2)$。**Du Fort-Frankel** 格式的相容性是有条件的，但这种相容性特点却表明了抛物型方程和双曲型方程之间的一种形式上的联系。

再考虑稳定性问题。仿照 **Richardson** 格式的稳定性分析过程，设$U_j^k=(u_j^k,v_j^k)^T$，则可将三层格式（1.41）改写为二层格式：

$$\begin{pmatrix} 1+2r & 0 \\ 0 & 1 \end{pmatrix} U_j^{k+1}$$

$$= \begin{pmatrix} 2r & 0 \\ 0 & 0 \end{pmatrix} U_{j-1}^k + \begin{pmatrix} 0 & 1-2r \\ 1 & 0 \end{pmatrix} U_j^k + \begin{pmatrix} 2r & 0 \\ 0 & 0 \end{pmatrix} U_{j+1}^k.$$

令$U_j^k = V^k(l) e^{i\sigma h}$，代入上式有

$$\begin{pmatrix} 1+2r & 0 \\ 0 & 1 \end{pmatrix} V^{k+1}(l) = \begin{pmatrix} 2r(e^{-i\sigma h} + e^{i\sigma h}) & 1-2r \\ 1 & 0 \end{pmatrix} V^k(l),$$

从而可算出传播矩阵

$$G(\sigma,\tau) = \begin{pmatrix} 2a\cos\sigma h & (1-a)/(1+a) \\ 1 & 0 \end{pmatrix},$$

其中，$a = 2r$。G的特征方程是

$$\mu^2 - \left(\frac{2a}{1+a}\cos\sigma h\right)\mu - \frac{1-a}{1+a} = 0,$$

其特征值为

$$\mu_{1,2} = \frac{a\cos\sigma h \pm \sqrt{1 - a^2\sin^2\sigma h}}{1+a},$$

利用下面即将给出的引理1.1，可知 $|\mu_{1,2}| \leqslant 1$，

若 $\mu_{1,2}$ 是重根，此时 $|\mu_{1,2}| \leqslant a/(1+a) < 1$，利用定理1.7中的2°，则证得此时差分格式稳定。

若 $\mu_{1,2}$ 不是重根，此时由 $|\mu_{1,2}| \leqslant 1$推得$\rho(G(\sigma,\tau)) \leqslant 1$，即满足 von Neumann 条件，利用定理 1.7 中的 1°，则证得此时差分格式也是稳定的。

总之，Du Fort-Frankel 格式是绝对稳定的。

引理1.1 实系数二次方程 $\mu^2 - b\mu - c = 0$ 的根 $|\mu_{1,2}| \leqslant 1$ 的充要条件是 $|b| \leqslant 1 - c$ 且 $|c| \leqslant 1$。

5.3 三层隐式格式

用时间方向上一阶向前和向后差商的加权平均修改古典隐格

式，就得到**三层隐式格式**：

$$\frac{3}{2}\frac{u_j^{k+1}-u_j^k}{\tau}-\frac{1}{2}\frac{u_j^k-u_j^{k-1}}{\tau}-a^2\frac{u_{j-1}^{k+1}-2u_j^{k+1}+u_{j+1}^{k+1}}{h^2}$$

$$=0.$$

易知其截断误差 $R_j^k=O(\tau^2+h^2)$.关于稳定性分析，可仿照 Du Fort-Frankel 格式的分析过程，先将上式化成等价的二层差分方程组，再求出传播矩阵

$$G(\sigma,\tau)=\begin{pmatrix}\dfrac{4}{3+8r\sin^2\dfrac{\sigma h}{2}} & \dfrac{-1}{3+8r\sin^2\dfrac{\sigma h}{2}}\\ 1 & 0\end{pmatrix},$$

$G(\sigma,\tau)$的特征方程是

$$\mu^2-\frac{4}{3+8r\sin^2\dfrac{\sigma h}{2}}\mu+\frac{1}{3+8r\sin^2\dfrac{\sigma h}{2}}=0,$$

利用引理1.1知 $|\mu_{1,2}|\leqslant 1$. 再把特征值具体写出：

$$\mu_{1,2}=\left(2\pm\sqrt{1-8r\sin^2\frac{\sigma h}{2}}\right)\Big/\left(3+8r\sin^2\frac{\sigma h}{2}\right).$$

若 $\mu_{1,2}$ 有重根，此时 $|\mu_{1,2}|<1$，则利用定理 1.7 中的2°；若 $\mu_{1,2}$ 相异，此时 $|\mu_{1,2}|\leqslant 1$，则利用定理 1.7 中的 1°；可证得三层隐式格式是绝对稳定的.

5.4 交替显隐格式

如前所述，我们将古典显格式和古典隐格式作加权平均，取得了较满意的效果.这里将介绍交替使用古典显、隐格式的效果.

跳点格式

把网格点(x_j,t_k)按 $k+j$ 等于偶数或等于奇数分成两组，分别称作**偶数网格点**和**奇数网格点**. 从t_k到 t_{k+1} 的计算过程中，先在偶数点处用古典显格式，再在奇数点处用古典隐格式计算，即

$$
\begin{cases}
u_j^{k+1}-u_j^k-r(u_{j-1}^k-2u_j^k+u_{j+1}^k)=0, \\
\qquad\qquad\qquad\qquad k+1+j=\text{偶}, & (1.42) \\
u_j^{k+1}-u_j^k-r(u_{j-1}^{k+1}-2u_j^{k+1}+u_{j+1}^{k+1})=0, \\
\qquad\qquad\qquad\qquad k+1+j=\text{奇}. & (1.43)
\end{cases}
$$

这一跳点格式的计算过程还有较简单的规则可循。当算出$k+1$层上全部节点值后，要用(1.42)式的相应公式

$$u_j^{k+2}-u_j^{k+1}-r(u_{j-1}^{k+1}-2u_j^{k+1}+u_{j+1}^{k+1})=0, \quad k+2+j=\text{偶},$$

计算偶数点处的值，因为$k+2$层上的偶数网格点和$k+1$层上的奇数网格点相同，所以上式与(1.43)式相减得$u_j^{k+2}=2u_j^{k+1}-u_j^k$，故$k+2$层上的节点值可由下式计算

$$
\begin{cases}
u_j^{k+2}=2u_j^{k+1}-u_j^k, \quad k+2+j=\text{偶}, \\
u_j^{k+2}-u_j^{k+1}-r(u_{j-1}^{k+2}-2u_j^{k+2}+u_{j+1}^{k+2})=0, \quad k+2+j=\text{奇}.
\end{cases}
\tag{1.44}
$$

因此，跳点格式的计算过程是这样的：当$k=1$时，用(1.42)式和(1.43)式计算；以后各层都可用(1.44)式计算。另外，当偶数点处的值显式算出后，奇数点值的计算也是显式计算，所以跳点格式的计算量很小。另外，跳点格式只需一套存贮单元。

值得注意的是，用(1.44)式的第一式和(1.43)式消去u_j^{k+1}，就有$k+1$层的奇数点或$k+2$层的偶数点上的计算式

$$\frac{1}{2\tau}(u_j^{k+2}-u_j^k)-\frac{a^2}{h^2}(u_{j-1}^{k+1}-u_j^{k+2}-u_j^k+u_{j+1}^{k+1})=0.$$

这是Du Fort-Frankel格式，其截断误差有二阶精度；又因为利用了二阶精度的偶数点值去显式计算奇数点值，所以跳点格式的相容性、稳定性和精度都同于Du Fort-Frankel格式，但计算方法上却有较大改善。

显式预测隐式校正格式

在$k+(1/2)$层上，用古典显格式计算出预测值（过渡值）$u_j^{k+1/2}$；再在$k+1$层上用古典隐格式校正预测值，算出校正值

u_j^{k+1}. 该格式的计算公式是

$$\begin{cases} \dfrac{u_j^{k+1/2}-u_j^k}{\tau/2} - a^2\dfrac{u_{j-1}^k - 2u_j^k + u_{j+1}^k}{h^2}=0, \\ \dfrac{u_j^{k+1}-u_j^{k+1/2}}{\tau/2} - a^2 \dfrac{u_{j-1}^{k+1}-u_j^{k+1}+u_{j+1}^{k+1}}{h^2}=0. \end{cases} \tag{1.45}$$

只要由上两式消去中间值$u_j^{k+1/2}$，即得到Crank-Nicholson格式，故格式(1.45)是截断误差为$O(\tau^2+h^2)$的绝对稳定格式。

隐式预测显式校正格式

也可采用另一种组合，即在$k+1/2$层上用古典隐格式计算出预测值$u_j^{k+1/2}$；再在$k+1$层上用古典显格式求出校正值 u_j^{k+1}. 该算法计算公式是

$$\begin{cases} \dfrac{u_j^{k+1/2}-u_j^k}{\tau/2} - a^2\dfrac{u_{j-1}^{k+1/2} - 2u_j^{k+1/2} + u_{j+1}^{k+1/2}}{h^2}=0, \\ \dfrac{u_j^{k+1}-u_j^{k+1/2}}{\tau/2} - a^2 \dfrac{u_{j-1}^{k+1/2}-2u_j^{k+1/2}+u_{j+1}^{k+1/2}}{h^2}=0. \end{cases} \tag{1.46}$$

只要由上两式消去中间值 $u_j^{k+1/2}$，即得到 Richardson 格式，所以格式 (1.46) 的截断误差 $R_j^k=O(\tau^2+h^2)$. 但此格式是绝对稳定的，事实上，上面第一式的传播因子是 $G_1=\left(1+4r\sin^2\dfrac{\sigma h}{2}\right)^{-1}$，第二式的传播因子是 $G_2=1-4r\sin^2\dfrac{\sigma h}{2}$，故(1.46)格式的传播因子$G=G_2\cdot G_1$，$|G|\leqslant 1$ (r任意).

最后，我们要指出，在数值求解一个空间变量的常系数扩散方程中，预测校正格式基本不用，这是因为 Crank-Nicholson 格式不仅精度好而且使用也方便的缘故。但是，这些交替使用显式和隐式的算法设计思想是很有启发性的；它对高维问题（空间变量个数$\geqslant 2$）的高效率算法设计有指导作用（参见§7）；另外，预测校正格式在非线性方程数值解中应用也很广泛，效果也好。

§6 守恒型差分格式与能量估计

前面§5中所讨论的诸种差分格式，都是就线性常系数方程采用直接差分化方法的结果。直接差分化方法也可推广到变系数方程的情形。例如

$$\left[\frac{\partial u}{\partial t}\right]_j^k=\left[a(x,t)\frac{\partial^2 u}{\partial x^2}+f(x,t)\right]_j^k$$

$$=a_j^k\left[\frac{\partial^2 u}{\partial x^2}\right]_j^k+f_j^k,$$

其中a_j^k和f_j^k分别表示$a(x_j,t_k)$和$f(x_j,t_k)$，用差商代替微商就得到一个差分格式。这种在节点处固定变系数数值后构造差分格式的方法称为**冻结系数法**；仍可用 Fourier方法分析相应格式的稳定性。冻结系数法简单易行而且直观，表现有点生硬是其不足之处。本节将介绍合理有效地处理变系数偏微分方程的积分插值法，并用能量估计分析差分格式的稳定性，该方法对变系数$a(x,t)$的大幅度变化和间断的情形也是适用的。

6.1 守恒型方程的积分插值法

因为很多物理现象或运动过程常常服从某种守恒规律，因而描述它们数量关系的微分方程也保持着这种守恒特征。本节考虑**变系数守恒型抛物型方程**的第一初边值问题：

$$\begin{cases}\dfrac{\partial u}{\partial t}-\dfrac{\partial}{\partial x}\left(p(x,t)\dfrac{\partial u}{\partial x}\right)=f(x,t), & (x,t)\in D,\\ u(x,0)=\varphi(x), & x\in[0,1],\\ u(0,t)=u(1,t)=0, & t\in[0,T],\end{cases}$$

其中，$\overline{D}=\{(x,t)\mid x\in[0,1],\ t\in[0,T]\}$，$p>0$。

仍和图 1.1 那样对 $\overline{D}$ 作网格剖分，并在小矩形域 $\Big\{(x,t)\mid$

$x_j-\frac{h}{2}\leqslant x\leqslant x_j+\frac{h}{2}$，$t_k\leqslant t\leqslant t_{k+1}\}$上对(1.46)式积分，有

$$\int_{t_k}^{t_{k+1}}\int_{x_j-h/2}^{x_j+h/2}\frac{\partial u}{\partial t}\,dxdt-\int_{t_k}^{t_{k+1}}\int_{x_j-h/2}^{x_j+h/2}\frac{\partial}{\partial x}\left(p\frac{\partial u}{\partial x}\right)dxdt$$

$$=\int_{t_k}^{t_{k+1}}\int_{x_j-h/2}^{x_j+h/2}f\,dxdt.$$

取$w=-p\partial u/\partial x$表示热流密度，得

$$\int_{x_j-h/2}^{x_j+h/2}[u(x,t_{k+1})-u(x,t_k)]dx$$

$$+\int_{t_k}^{t_{k+1}}[w(x_{j+\frac{1}{2}},t)-w(x_{j-\frac{1}{2}},t)]dt$$

$$=\int_{t_k}^{t_{k+1}}\int_{x_j-h/2}^{x_j+h/2}f\,dxdt, \tag{1.47}$$

这个积分式在物理上反应了热量的平衡．再利用矩形积分公式，有

$$\int_{x_j-h/2}^{x_j+h/2}[u(x,t_{k+1})-u(x,t_k)]dx$$

$$=h[u(x_j,t_{k+1})-u(x_j,t_k)]+O(h^3),$$

利用梯形公式有

$$\int_{t_k}^{t_{k+1}}[w(x_{j+\frac{1}{2}},t)-w(x_{j-\frac{1}{2}},t)]dt$$

$$=\frac{\tau}{2}[w(x_{j+\frac{1}{2}},t_{k+1})-w(x_{j-\frac{1}{2}},t_{k+1})+w(x_{j+\frac{1}{2}},t_k)$$

$$-w(x_{j-\frac{1}{2}},t_k)]+O(\tau^3)$$

$$=\theta\tau[w(x_{j+\frac{1}{2}},t_{k+1})-w(x_{j-\frac{1}{2}},t_{k+1})]$$

$$+(1-\theta)\tau[w(x_{j+\frac{1}{2}},t_k)-w(x_{j-\frac{1}{2}},t_k)]$$

$$+O\left[\left(\frac{1}{2}-\theta\right)\tau^2+\tau^3\right], \tag{1.48}$$

其中$0\leqslant\theta\leqslant1$．由于$-\partial u/\partial x=w/p$，积分后有

$$u(x_{j-1},t)-u(x_j,t)=\int_{x_{j-1}}^{x_j}\frac{w}{p}dx$$

$$=w(x_{j-\frac{1}{2}},t)\int_{x_{j-1}}^{x_j}\frac{1}{p}dx+O(h^3).$$

记$A_j^t=h\left[\int_{x_{j-1}}^{x_j}\frac{1}{p}dx\right]^{-1}$，代人上式得

$$A_j^t\frac{u(x_{j-1},t)-u(x_j,t)}{h}=w(x_{j-\frac{1}{2}},t)+O(h^3),$$

同理有

$$A_{j+1}^t\frac{u(x_j,t)-u(x_{j+1},t)}{h}=w(x_{j+\frac{1}{2}},t)+O(h^3).$$

将这两个式子代入(1.48)式，并记$A_j^{t_k}=A_j^k$，则得到

$$\int_{t_k}^{t_{k+1}}[w(x_{j+\frac{1}{2}},t)-w(x_{j-\frac{1}{2}},t)]dt$$

$$=\frac{\theta\tau}{h}\{A_{j+1}^{k+1}[u(x_j,t_{k+1})-u(x_{j+1},t_{k+1})]$$

$$-A_j^{k+1}[u(x_{j-1},t_{k+1})-u(x_j,t_{k+1})]\}$$

$$+\frac{(1-\theta)\tau}{h}\{A_{j+1}^k[u(x_j,t_k)-u(x_{j+1},t_k)]$$

$$-A_j^k[u(x_{j-1},t_k)-u(x_j,t_k)]\}$$

$$+O\left[\left(\frac{1}{2}-\theta\right)\tau^2+\tau^3+h^3\right].$$

应用差商符号：

$$\Delta u(x_j,t_k)=u(x_{j+1},t_k)-u(x_j,t_k),$$

$$\nabla u(x_j,t_k)=u(x_j,t_k)-u(x_{j-1},t_k),$$

可简化上式，得

$$\int_{t_k}^{t_{k+1}}[w(x_{j+\frac{1}{2}},t)-w(x_{j-\frac{1}{2}},t)]dt$$

$$= -\frac{\theta\tau}{h}[\Delta(A_j^{k+1}\nabla u(x_j,t_{k+1}))] - \frac{(1-\theta)\tau}{h}$$

$$\cdot[\Delta(A_j^k\nabla u(x_j,t_k))] + O\left[\left(\frac{1}{2}-\theta\right)\tau^2+\tau^3+h^3\right].$$

这样，积分式(1.47)就变成

$$h[u(x_j,t_{k+1})-u(x_j,t_k)] - \frac{\theta\tau}{h}[\Delta(A_j^{k+1}\nabla u(x_j,t_{k+1}))]$$

$$-\frac{(1-\theta)}{h}\tau[\Delta(A_j^k\nabla u(x_j,t_k))] = \int_{x_{j-\frac{1}{2}}}^{x_{j+\frac{1}{2}}}\int_{t_k}^{t_{k+1}} f dx dt$$

$$+O\left[\left(\frac{1}{2}-\theta\right)\tau^2+\tau^3+h^3\right].$$

从而得差分方程

$$\frac{u_j^{k+1}-u_j^k}{\tau} - \frac{1}{h^2}[\theta\Delta(A_j^{k+1}\nabla u_j^{k+1}) + (1-\theta)\Delta(A_j^k\nabla u_j^k)]$$

$$= f_j^k, \tag{1.49}$$

式中$f_j^k = \frac{1}{\tau h}\int_{x_{j-\frac{1}{2}}}^{x_{j+\frac{1}{2}}}\int_{t_k}^{t_{k+1}} f(x,t)dxdt$。(1.49)式是加权六点格式；当$\theta=0$时得古典显格式；$\theta=1$时得古典隐格式；$\theta=\frac{1}{2}$时得Crank-Nicholson格式。

要使用(1.49)式，必须利用数值积分计算出A_j^k和f_j^k，但要注意数值积分误差阶不低于截断误差的阶数，否则会降低差分解的精度。

按上述方法构造的差分格式(1.49)近似地满足热量守恒条件，故称为**守恒型格式**。这种格式是利用积分形式并通过插值导出的，因此这种构造方法又叫**积分插值法**。积分插值法除能保持研究对象的某些物理特征和数学特征外，还能适应不同光滑程度的变系数以及不均匀步长情形。采用该方法时，对间断点的处理以及对各种边值条件的处理也是方便的，有兴趣的读者可参见

[1-4].

6.2 能量估计

前面介绍的 Fourier 分析法，只能有效地用于判别常系数线性方程的差分格式的稳定性，对变系数情形尚在发展中。这里，介绍一种应用范围较广泛的**能量估计方法**，该方法是研究稳定性和收敛性的最一般方法，它是微分方程理论研究中**能量方法**的模拟。但这种方法不能象 Fourier 分析法那样对某类格式建立统一的判别法则，只能具体处理和估计。

为了不使繁琐的演算掩盖能量估计方法的本质，假定(1.46)式中的 p 与 t 无关，即 $A_j^k = A_j$，讨论(1.49)式中 $\theta=1$ 时的古典隐格式

$$\begin{cases} u_j^k - u_j^{k-1} - r\Delta(A_j \nabla u_j^k) = \tau f_j^{k-1}, \\ u_j^0 = \varphi(x_j), \qquad u_0^k = u_N^k = 0. \end{cases} \tag{1.50}$$

的稳定性。

首先，设 $\varphi(x)$ 和 $\psi(x)$ 是任意两个满足齐次边界条件的函数，因为

$$\varphi_0 = \varphi_N = \psi_0 = \psi_N = 0,$$

$$\Delta(\varphi_j, \psi_j) = \varphi_{j+1}\psi_{j+1} - \varphi_j\psi_j = \varphi_{j+1}\psi_{j+1} - \varphi_j\psi_{j+1} + \varphi_j\psi_{j+1} - \varphi_j\psi_j = \psi_{j+1}\Delta\varphi_j + \varphi_j\Delta\psi_j,$$

$$\sum_{j=0}^{N-1}\Delta(\varphi_j, \psi_j) = \varphi_N\psi_N - \varphi_0\psi_0 = 0,$$

所以

$$\sum_{j=0}^{N-1}\psi_{j+1}\Delta\varphi_j + \sum_{j=0}^{N-1}\varphi_j\Delta\psi_j = 0,$$

利用 $\Delta\varphi_j = \nabla\varphi_{j+1}$，得到**Abel分部求和公式**

$$\sum_{j=0}^{N-1}\varphi_j\Delta\psi_j = -\sum_{j=1}^{N}\psi_j\nabla\varphi_j. \tag{1.51}$$

再记 $\delta_t u_j^k = u_j^k - u_j^{k-1}$，将其乘（1.50）式两端并对 j 从 1 到 $N-1$ 求和，还利用(1.51)式，就有

$$\sum_{j=1}^{N-1}(\partial_t u_j^k)^2 - r\sum_{j=1}^{N-1}\partial_t u_j^k\Delta(A_j\nabla u_j^k) = \tau\sum_{j=1}^{N-1}f_j^{k-1}\partial_t u_j^k,$$

$$\sum_{j=1}^{N-1}(\partial_t u_j^k)^2 + r\sum_{j=1}^{N}A_j\nabla u_j^k\nabla(\partial_t u_j^k) = \tau\sum_{j=1}^{N-1}f_j^{k-1}\partial_t u_j^k.$$

将恒等式

$$u_j^k = \frac{1}{2}(u_j^k + u_j^{k-1}) + \frac{1}{2}\partial_t u_j^k$$

代入上式整理得

$$\sum_{j=1}^{N-1}(\partial_t u_j^k)^2 + \frac{r}{2}\sum_{j=1}^{N}A_j(\nabla u_j^k)^2 + \frac{r}{2}\sum_{j=1}^{N}A_j[\nabla(\partial_t u_j^k)]^2$$

$$= \tau\sum_{j=1}^{N-1}f_j^{k-1}\partial_t u_j^k + \frac{r}{2}\sum_{j=1}^{N}A_k(\nabla u_j^{k-1})^2.$$

利用不等式$ab\leqslant\varepsilon a^2+b^2/4\varepsilon$（$a$、$b$、$\varepsilon$为任意实数，$\varepsilon>0$），导出

$$\tau\sum_{j=1}^{N-1}f_j^{k-1}(\partial_t u_j^k) \leqslant \frac{\tau}{4\varepsilon}\sum_{j=1}^{N-1}(f_j^{k-1})^2 + \varepsilon\tau\sum_{j=1}^{N-1}(\partial_t u_j^k)^2.$$

将此式（选$\varepsilon\tau=1$）代入上式并舍去左端第三项，得到

$$\frac{r}{2}\sum_{j=1}^{N}A_j(\nabla u_j^k)^2 \leqslant \frac{r}{2}\sum_{j=1}^{N}A_j(\nabla u_j^{k-1})^2 + \frac{\tau^2}{4}\sum_{j=1}^{N-1}(f_j^{k-1})^2.$$

记　$I_k = h\sum_{j=1}^{N}A_j\left(\frac{\nabla u_j^k}{h}\right)^2$，则上式可改写为

$$I_k \leqslant I_{k-1} + \frac{\tau}{2}\sum_{j=1}^{N-1}(f_j^{k-1})^2 h.$$

取　$\|f^k\|^2 = h\sum_{j=1}^{N-1}[f_j^k]^2$，就有

$$I_k \leqslant I_{k-1} + \frac{\tau}{2}\|f^{k-1}\|^2.$$

反复使用上式，最后得到

$$I_k \leqslant I_0 + \frac{\tau}{2}\sum_{j=1}^{N-1}\|f^l\|^2.$$

从物理角度看，由于 $I_k = \|u_h^k\|_1^2 = h\sum_{j=1}^{N} A_j\left(\frac{\nabla u_j^k}{h}\right)^2$ 模拟了 $\|u(x,\ t_k)\|_1^2 = \int_0^1 A\left(\frac{\partial u(x,t_k)}{\partial x}\right)^2 dx$，故上式称作差分方程(1.50)的能量估计，它说明差分方程(1.50)按 $\|\cdot\|_1$ 绝对稳定。

与上面的推导相同，可证格式(1.50) 当 $\theta \geqslant \frac{1}{2}$ 时绝对稳定；也同样可以求得 $\theta < \frac{1}{2}$ 时的稳定条件，特别当 $\theta = 0$ 时按 $\frac{\tau}{h^2} \leqslant 1/(2\max|A_j|)$ 条件稳定；若 A_j 恒为常数，则这些结果与 Fourier 分析法结果一致。

§7　解二维问题的分裂算法

7.1　二维问题差分方法的一般性和特殊性

前面讨论了一个空间变量(一维)抛物型方程的差分方法，那里介绍的一切概念、差分格式构造办法、稳定性分析法和有关结论都可以简单地推广到多维情形。现以二维模型问题

$$\begin{cases} \frac{\partial u}{\partial t} - a^2\left(\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2}\right) = 0, & (x,y,t) \in (0,1)^2 \times (0,T), \\ u(x,y,0) = \varphi(x,y) & (0 \leqslant x、\ y \leqslant 1), \\ u(0,y,t) = u(1,y,t) = u(x,0,t) = u(x,1,t) = 0, & 0 \leqslant t \leqslant T \end{cases} \tag{1.52}$$

为例来说明这种推广的简单性。相应的区域分划参数记为 h_1、h_2、τ，还记 $r_1 = a^2\tau/h_1^2$，$r_2 = a^2\tau/h_2^2$。

对二维常系数线性方程，可仿照一维的相应情形，采用直接差分化方法可得到对应于§5中的各种差分格式，也可用 Fourier 分析方法分析其稳定性。

例如，用一阶向前差商代替 $\partial u/\partial t$，用二阶中心差商

$$\delta_x^2 u_{i,j}^k/h_1^2 = (u_{i-1,j}^k - 2u_{i,j}^k + u_{i+1,j}^k)/h_1^2,$$

$$\delta_y^2 u_{ij}^k/h_2^2=(u_{i,j-1}^k-2u_{i,j}^k+u_{i,j+1}^k)/h_2^2,$$

分别代替$\partial^2 u/\partial x^2$和$\partial^2 u/\partial y^2$，则得到古典显格式

$$\frac{1}{\tau}(u_{ij}^{k+1}-u_{ij}^k)=a^2(\delta_x^2 u_{ij}^k/h_1^2+\delta_y^2 u_{ij}^k/h_2^2).$$

将u_{ij}^k扩展定义为k时间层上的阶梯函数$u_h^k(x_1,x_2)$并开拓到整个x_1-x_2平面上，就可将古典显格式改写为

$$\begin{aligned}u_h^{k+1}(x_1,x_2)=&(1-2r_1-2r_2)u_h^k(x_1,x_2)\\&+r_1(u_h^k(x_1-h_1,x_2)+u_h^k(x_1+h_1,x_2))\\&+r_2(u_h^k(x_1,x_2+h)+u_h^k(x_1,x_2-h)).\end{aligned}$$

记$x=(x_1,x_2)$，$h=(h_1,h_2)$，$\sigma=(\sigma_1,\sigma_2)=(\pi l_1,\pi l_2)$仿照一维Fourier分析法过程，令$u_h^k(x)=v^k(l)$. $e^{i\sigma x}$代入上式，有

$$\begin{aligned}e^{i\sigma x}v^{k+1}(l)=[&(1-2r_1-2r_2)+r_1(e^{i\sigma_1 h_1}+e^{-i\sigma_1 h_1})\\&+r_2(e^{i\sigma_2 h_2}+e^{-i\sigma_2 h_2})]e^{i\sigma x}v^k(l),\end{aligned}$$

消去公因子$e^{i\sigma x}$，就得到$v^{k+1}(l)=Gv^k(x)$，其中传播矩阵

$$G=1-4r_1\sin^2\frac{\sigma_1 h_1}{2}-4r_2\sin^2\frac{\sigma_2 h_2}{2}.$$

令$r=r_1=r_2=$常数，因Von-Neumann条件$|G|\leqslant 1$是古典显格式稳定的充要条件，所以由上式可推出$r\leqslant 1/4$是古典显格式稳定性条件。

对二维变系数问题，例如(1.52)式中$a=a(x,t)$，一种简单的处理办法是**冻结系数法**。(1.52)式采用冻结系数处理所得到的古典隐格式可写为

$$\frac{1}{\tau}(u_{ij}^{k+1}-u_{ij}^k)=(a_{ij}^k)^2(\delta_x^2 u_{ij}^{k+1}+\delta_y^2 u_{ij}^{k+1})/h^2.$$

但是，用 Von-Neumann 条件判别上式绝对稳定却只能说明(x_i,y_j,t_k)附近的情形，还要就全部的i，j加以考虑，才能说明全区域中差分格式的稳定性。

对于二维变系数问题，积分插值法和能量估计也是普遍适用的。

虽然，二维问题的差分方法可由一维情形简单推广得到，但是它也有自己的特殊性。

二维显格式计算简单，但稳定性条件比一维相应情形要苛刻得多。如一维显格式的稳定性条件是$r\leqslant 1/2$，而n维显格式的稳定性条件将是$r\leqslant 1/2n$。因此，要维持多维显格式的稳定性需付出较大的计算代价。

二维隐格式最大优点是绝对稳定。在实际计算中，步长选取不受限制，这是非常方便而实用的。例如，二维Crank-Nicholson格式是一个截断误差为$O(\tau^2+h^2)$的绝对稳定格式，是很实用的。然而应该看到，解二维隐格式一般要求解5对角方程组，其计算量比求解一维隐格式所对应的3对角方程组（可用追赶法快速求解）的计算量要大很多。维数增高，这一现象更加突出。

因此，对高维问题，构造绝对稳定且计算量又小的差分格式一直是抛物型方程数值解法中重要的研究课题。

7.2 分裂算法的基本思想

由一维交替显隐格式可知，几个简单格式可以复合成一个复杂算法。由此得到启发，若能将高维差分格式分解为几个可简单求解的低维（特别是一维）差分格式，则必能有效地降低高维差分格式的求解计算量。

先观察解(1.52)式的 Crank-Nicholson 格式

$$\frac{u_{ij}^{k+1}-u_{ij}^{k}}{\tau}=\frac{a^2}{2h^2}[(\delta_x^2u_{ij}^{k+1}+\delta_y^2u_{ij}^{k+1})+(\delta_x^2u_{ij}^{k}+\delta_y^2u_{ij}^{k})],$$

将其变形为

$$u_{ij}^{k+1}-\frac{r}{2}\delta_x^2u_{ij}^{k+1}-\frac{r}{2}\delta_y^2u_{ij}^{k+1}=u_{ij}^{k}+\frac{r}{2}\delta_x^2u_{ij}^{k}+\frac{r}{2}\delta_y^2u_{ij}^{k},$$

两边分别加上$\frac{1}{4}r^2\delta_x^2\delta_y^2u_{ij}^{k+1}$和$\frac{1}{4}r^2\delta_x^2\delta_y^2u_{ij}^{k}$，且分解因式，就有

$$\left(1-\frac{r}{2}\delta_x^2\right)\left(1-\frac{r}{2}\delta_y^2\right)u_{ij}^{k+1}$$

$$=\left(1+\frac{r}{2}\delta_x^2\right)\left(1+\frac{r}{2}\delta_y^2\right)u_{ij}^{k}. \qquad (1.53)$$

先分析 (1.53) 式的相容性和截断误差。因用于修改Crank-Nicholson 格式的添加项

$$\frac{r^2}{4}\delta_x^2\delta_y^2(u_{ij}^{k+1}-u_{ij}^{k})=\frac{\tau^2}{4}[u_{xxyy}]_{ij}^{k}+O(\tau^3+h^4),$$

所以(1.53)格式当 $\tau\to0, h\to0, h=g_r(\tau)$ 时，仍相容于(1.52)式中的微分方程；并且格式(1.53)完全相同于 Crank-Nicholson 格式的相容性和截断误差。值得注意的是，由于添补项的原因，格式(1.53)要求方程(1.52)的解有较高的光滑性，即 $u\in C^2[0,T]\times C^4([0,1]\times[0,1])$；另外，由上式可知，添补项的效果相当于原方程中引进了**粘性项** $(\partial^4u/\partial x^2\partial y^2)\tau^2/4$，它非但不影响相容性，而且有助差分格式的简化，这也是抛物型和双曲型方程差分方法中常常使用的办法之一。

再分析(1.53)式的稳定性。用Fourier 分析法很容易得到相应的传播矩阵

$$G(\sigma,\tau)=\frac{\left(1-2r\sin^2\frac{\sigma_1h}{2}\right)\left(1-2r\sin^2\frac{\sigma_2h}{2}\right)}{\left(1+2r\sin^2\frac{\sigma_1h}{2}\right)\left(1+2r\sin^2\frac{\sigma_2h}{2}\right)},$$

利用 Von-Neumann 条件知，(1.53)格式绝对稳定。

(1.53) 式能分解成两个简单的一维格式，下面介绍可由(1.53)式生成的几种**分裂算法**。

7.3 P-R (Peaceman-Rachford)格式

将(1.53)式分解为

$$\begin{cases}\left(1-\dfrac{r}{2}\delta_x^2\right)u_{ij}^{k+1/2}=\left(1+\dfrac{r}{2}\delta_y^2\right)u_{ij}^{k}, \\ \left(1-\dfrac{r}{2}\delta_y^2\right)u_{ij}^{k+1}=\left(1+\dfrac{r}{2}\delta_x^2\right)u_{ij}^{k+1/2},\end{cases}$$

这就是 **Peaceman-Rachford 格式** (1955年)，简称为 **P-R 格式**。

P-R 格式消去中间变量 $u_{ij}^{k+1/2}$ 就得(1.53)式，故 P-R 格式是截断误差为 $O(\tau^2+h^2)$ 的绝对稳定格式。

P-R 格式是分别在 x 和 y 方向交替使用一维隐显格式的结果，每时间层只需解两个 3 对角方程组，所以计算量小，只相当于 1/7 的解二维问题的Crank-Nicholson 格式的计算量。P-R 格式形式对称，结构简单合理，且只需一套存贮单元。不过P-R 格式不能推广到三维情形。

7.4 Douglas 格式

将(1.53)式变形为

$$\left(1-\frac{r}{2}\delta_x^2\right)\left(1-\frac{r}{2}\delta_y^2\right)u_{ij}^{k+1}$$
$$=\left[\left(1-\frac{r}{2}\delta_x^2\right)\left(1-\frac{r}{2}\delta_y^2\right)+r(\delta_x^2+\delta_y^2)\right]u_{ij}^{k},$$

$$\left(1-\frac{r}{2}\delta_x^2\right)\left(1-\frac{r}{2}\delta_y^2\right)\frac{u_{ij}^{k+1}-u_{ij}^{k}}{\tau}$$
$$=\frac{a^2}{h^2}(\delta_x^2+\delta_y^2)u_{ij}^{k}.$$

上式可分裂为

$$\begin{cases}\left(1-\dfrac{r}{2}\delta_x^2\right)(u_{ij}^{k+1/2}-u_{ij}^{k})=r(\delta_x^2+\delta_y^2)u_{ij}^{k}, \\ \left(1-\dfrac{r}{2}\delta_y^2\right)(u_{ij}^{k+1}-u_{ij}^{k})=u_{ij}^{k+1/2}-u_{ij}^{k}.\end{cases}$$

整理得 **Douglas 格式**：

$$
\begin{cases}
\left(1-\dfrac{r}{2}\delta_x^2\right)\left(u_{ij}^{k+1/2}-u_{ij}^k\right)=r\left(\delta_x^2+\delta_y^2\right)u_{ij}^k, \\
\dfrac{2}{\tau}\left(u_{ij}^{k+1}-u_{ij}^{k+1/2}\right)=\dfrac{a^2}{h^2}\delta_y^2\left(u_{ij}^{k+1}-u_{ij}^k\right).
\end{cases}
$$

Douglas 格式也是分别在 x 和 y 方向交替使用一维隐显格式的结果，它的计算量、截断误差和稳定性都同于P-R格式。因该格式的第二式中同时出现了u_{ij}^k、$u_{ij}^{k+1/2}$和u_{ij}^{k+1}，故Douglas格式需要两套存贮单元。另外，Douglas格式还可推广到三维情形。

例如三维模型问题

$$
\frac{\partial u}{\partial t}=a^2\left(\frac{\partial^2 u}{\partial x^2}+\frac{\partial^2 u}{\partial y^2}+\frac{\partial^2 u}{\partial z^2}\right)+f(x,y,z,t)
$$

的Douglas格式为

$$
\begin{cases}
\left(1-\dfrac{r}{2}\delta_x^2\right)\left(u_{ijm}^{k+1/3}-u_{ijm}^k\right)=r\left(\delta_x^2+\delta_y^2+\delta_z^2\right)u_{ijm}^k+\tau f_{ijm}^{k+1/2}, \\
\dfrac{2}{\tau}\left(u_{ijm}^{k+2/3}-u_{ijm}^{k+1/3}\right)=\dfrac{a^2}{h^2}\delta_y^2\left(u_{ijm}^{k+2/3}-u_{ijm}^k\right), \\
\dfrac{2}{\tau}\left(u_{ijm}^{k+1}-u_{ijm}^{k+2/3}\right)=\dfrac{a^2}{h^2}\delta_z^2\left(u_{ijm}^{k+1}-u_{ijm}^k\right).
\end{cases}
$$

7.5 Yanenko 格式

$$
\begin{cases}
\dfrac{u_{ij}^{k+1/4}-u_{ij}^k}{\tau/2}=\dfrac{a^2}{h^2}\delta_x^2 u_{ij}^{k+1/4}, \\
\dfrac{u_{ij}^{k+1/2}-u_{ij}^{k+1/4}}{\tau/2}=\dfrac{a^2}{h^2}\delta_y^2 u_{ij}^{k+1/2}, \\
\dfrac{u_{ij}^{k+1}-u_{ij}^k}{\tau}=\dfrac{a^2}{h^2}\left(\delta_x^2+\delta_y^2\right)u_{ij}^{k+1/2},
\end{cases}
$$

这就是Yanenko格式。由上式消去过渡值后，可知该格式与Douglas格式等价。此格式与Douglas格式有相同的计算特征，但计算公式稍简单些，它也可推广到三维情形。

7.6 局部一维格式

将(1.53)式分裂，就得到局部一维格式：

$$\begin{cases}\left(1-\dfrac{r}{2}\delta_x^2\right)u_{ij}^{k+1/2}=\left(1+\dfrac{r}{2}\delta_x^2\right)u_{ij}^k,\\ \left(1-\dfrac{r}{2}\delta_y^2\right)u_{ij}^{k+1}=\left(1+\dfrac{r}{2}\delta_y^2\right)u_{ij}^{k+1/2},\end{cases}$$

此格式除了有前面几种分裂算法的计算特征外，还有直接离散方程

$$\frac{1}{2}\ \frac{\partial u}{\partial t}=a^2\frac{\partial^2 u}{\partial x^2}\quad 和\quad \frac{1}{2}\ \frac{\partial u}{\partial t}=a^2\frac{\partial^2 u}{\partial y^2}$$

的特征。因上述两个方程相加就描述了$\frac{\partial u}{\partial t}=a^2\left(\frac{\partial^2 u}{\partial x^2}+\frac{\partial^2 u}{\partial y^2}\right)$，故局部一维格式表明了分裂算法的另一种构造办法，即先分裂微分方程再差分化。局部一维格式只需一套存贮单元，它也可以推广到三维情形。

§8 解非线性抛物型方程的差分方法

本节主要讨论求解非线性抛物型方程的差分方法，重点讨论几个重要的非线性抛物型方程的特殊解法。

为说明问题方便起见，考虑一个空间自变量的非线性抛物型方程的初边值问题：

$$\begin{cases}\dfrac{\partial u}{\partial t}=f\left(x,t,u,\dfrac{\partial u}{\partial x},\dfrac{\partial^2 u}{\partial x^2}\right), & x\in(0,1),\ t>0,\\ u(x,0)=v(x), & x\in[0,1],\\ u(0,t)=\psi_0(t), & t>0,\\ u(1,t)=\psi_1(t), & t\geqslant 0.\end{cases}\tag{1.54}$$

其中$\partial f\Big/\partial\left(\frac{\partial^2 u}{\partial x^2}\right)\geqslant a>0$。当然，人们可以把数值求解常系数

线性抛物型方程的许多方法和技巧推广到非线性情形。

对于非线性方程(1.54)，构造其差分格式的最简单的办法是采用显格式，即

$$\frac{u_j^{k+1}-u_j^k}{\tau}=f\left(x_j,t_k,u_j^k,\frac{u_{j+1}^k-u_{j-1}^k}{2h},\ \frac{u_{j+1}^k-2u_j^k+u_{j-1}^k}{h^2}\right),$$

$$j=1,2,\cdots,N-1. \tag{1.55}$$

很清楚，非线性差分格式(1.55)的稳定性分析复杂，且显格式稳定性条件限制较严。

还可采用稳定性好的隐格式，例如用Crank-Nicholson 格式：

$$\frac{u_j^{k+1}-u_j^k}{\tau}=f\left(x_j,t_{k+1/2},\quad \frac{1}{2}(u_j^k+u_j^{k+1}),\right.$$

$$\frac{1}{2}\left(\frac{u_{j+1}^k-u_{j-1}^k}{2h}+\frac{u_{j+1}^{k+1}-u_{j-1}^{k+1}}{2h}\right),$$

$$\left.\frac{1}{2h^2}\delta_x^2(u_j^k+u_j^{k+1})\right),\qquad j=1,2,\cdots,N-1, \tag{1.56}$$

其中，$t_{k+1/2}=\frac{1}{2}(t_k+t_{k+1})$。(1.56)式一般是稳定的，但求解$u_j^{k+1}$的代数方程组是非线性的，计算每一个时间层上的数值解都得求解一个相应的非线性代数方程组，都得付出较大的计算代价。

因此，人们自然期望在采用隐式计算时能简化非线性差分格式，减少计算工作量，增进数值求解速度。下面将围绕这个问题就几个特殊类型的非线性方程介绍几种有效的数值方法。

8.1 Richtmyer 线性化方法

考虑非线性抛物型方程

$$\frac{\partial u}{\partial t}=\frac{\partial^2(u^m)}{\partial x^2},\qquad m\geqslant 2. \tag{1.57}$$

Richtmyer 采用加权隐格式

$$\frac{u_j^{k+1}-u_j^k}{\tau}=\theta\frac{\delta_x^2(u^m)_j^{k+1}}{h^2}+(1-\theta)\frac{\delta_x^2(u^m)_j^k}{h^2} \tag{1.58}$$

逼近(1.57)式。

(1.58)式中的u_j^{k+1}是非线性的，将其 Taylor 展开，有

$$\begin{aligned}(u^m)_j^{k+1}&=(u^m)_j^k+\tau\left[\frac{\partial(u^m)}{\partial t}\right]_j^k+\cdots\\&=(u^m)_j^k+\tau\left[\frac{\partial(u^m)}{\partial u}\right]_j^k\left[\frac{\partial u}{\partial t}\right]_j^k+\cdots,\end{aligned}$$

Richtmyer 取其线性化部分，即有

$$(u^m)_j^{k+1}=(u^m)_j^k+m(u^{m-1})_j^k(u_j^{k+1}-u_j^k). \tag{1.59}$$

将(1.59)式代入(1.58)式且令$w_j=u_j^{k+1}-u_j^k$，就有

$$\frac{1}{\tau}w_j=\frac{\theta}{h^2}\delta_x^2[(u^m)_j^k+m(u^{m-1})_j^kw_j]+\frac{(1-\theta)}{h^2}\delta_x^2(u^m)_j^k.$$

整理得

$$\begin{aligned}\frac{1}{\tau}w_j&=\frac{m\theta}{h^2}\delta_x^2(u^{m-1})_j^kw_j+\frac{1}{h^2}\delta_x^2(u^m)_j^k\\&=\frac{m\theta}{h^2}[(u^{m-1})_{j-1}^kw_{j-1}-2(u^{m-1})_j^kw_j+(u^{m-1})_{j+1}^kw_{j+1}]\\&\quad+\frac{1}{h^2}[(u^m)_{j-1}^k-2(u^m)_j^k+(u^m)_{j+1}^k].\end{aligned} \tag{1.60}$$

若再考虑到初边值条件

$$w_0=u_0^{k+1}-u_0^k,\quad w_N=u_N^{k+1}-u_N^k,\qquad u_j^0=v_j,\qquad r=\frac{\tau}{h^2},$$

则得到一个关于变量 w_j 的 3 对角线性代数方程组，其矩阵形式为

$$\begin{pmatrix}1+2EF_1^k & -EF_2^k & & \\ -EF_1^k & 1+2EF_2^k & \ddots & \\ & \ddots & \ddots & -EF_{N-1}^k\\ & & -EF_{N-2}^k & 1+2EF_{N-1}^k\end{pmatrix}\begin{pmatrix}w_1\\ w_2\\ \vdots\\ w_{N-1}\end{pmatrix}$$

$$
=\begin{pmatrix} -2rF_1^k & rF_2^k & & \\ rF_1^k & -2rF_2^k & \ddots & \\ & \ddots & \ddots & rF_{N-1}^k \\ & & rF_{N-2}^k & -2rF_{N-1}^k \end{pmatrix}\begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_{N-1} \end{pmatrix}
$$

$$
+\begin{pmatrix} r(u^m)_0^k + EF_0^k(u_0^{k+1}-u_0^k) \\ 0 \\ \vdots \\ 0 \\ r(u^m)_N^k + EF_N^k(u_N^{k+1}-u_N^k) \end{pmatrix}
$$

其中，$E=mr\theta$，$F=u^{m-1}$。

所以，可逐层地计算出$w_j=u_j^{k+1}-u_j^k$，从而算出u_j^k只需花费解一个3对角线性方程组（可用追赶法求解）的代价，这是理想的。

(1.60)式不是常系数线性差分方程，只能采用“冻结系数”的“局部稳定性”分析．假设(1.60)式的系数$(u^{m-1})_j^k$看作“固定”的常数，它不再随j和k而变（这当然是在局部近似的意义下成立），令$m(u^{m-1})_j^k=a$，则(1.60)式简写为

$$
-r\theta a(u_{j-1}^{k+1}-u_{j-1}^k)+(1+2ar\theta)(u_j^{k+1}-u_j^k)-ra\theta(u_{j+1}^{k+1}-u_{j+1}^k)
$$

$$
=ar(u_{j+1}^k-2u_j^k+u_{j-1}^k),
$$

用 **Fourier** 分析法求得传播因子

$$
G(\sigma,\tau)=\frac{1-4(1-\theta)ar\sin^2\dfrac{\sigma h}{2}}{1+4ar\theta\sin^2\dfrac{\sigma h}{2}},
$$

从而求得差分格式(1.60)的局部稳定性条件是：

$$
\begin{cases} \text{当 } 0\leqslant\theta<1/2 \text{ 且 } ar\leqslant 1/2(1-2\theta)\text{ 时，格式稳定。} \\ \text{当 } \theta\geqslant 1/2 \text{ 时，格式绝对稳定。} \end{cases}
$$

作为计算实例，用 Richtmyer 方法求解定解问题：

$$\begin{cases} \dfrac{\partial u}{\partial t} = \dfrac{\partial^2 (u^2)}{\partial x^2}, & x \in (0,1) \\ u(x,0) = \sin \pi x, & x \in [0,1], \\ u(0,t) = u(1,t) = 0, & t \in [0,T]. \end{cases}$$

此问题的解为

$$(2u-3) + \ln\left(u - \frac{1}{2}\right) = 2(2t - x).$$

在数值计算中取$h = 0.1$，$r = \tau/h^2 = 0.5$，当$t = 0.5$时，用Richtmyer方法计算出如下满意的结果：

表1.3

x	准　确　解	近　似　解
0.1	2.149703	2.149701
0.3	1.997951	1.997948
0.5	1.849962	1.849958
0.7	1.706244	1.706240
0.9	1.567391	1.567389

8.2 Lees三层差分格式

Lees研究了非线性方程

$$b(u)\frac{\partial u}{\partial t} = \frac{\partial}{\partial x}\left(a(u)\frac{\partial u}{\partial x}\right), \quad b(u) > 0, \ a(u) > 0, \tag{1.61}$$

并用中心差商代替(1.61)式的微商，就有

$$b(u_j^k)\frac{u_j^{k+1} - u_j^{k-1}}{2\tau}$$

$$= \frac{1}{h}\left[a(u_{j+1/2}^k)\frac{u_{j+1}^k - u_j^k}{h} - a(u_{j-1/2}^k)\frac{u_j^k - u_{j-1}^k}{h}\right]. \tag{1.62}$$

显然，如果 $b(u) = a(u) = 1$，此格式变为 Richardson 格式，是不稳定的。为克服这一缺点，Lees作了下述处理：令

$$u_{j+1}^{k}\approx\frac{1}{3}(u_{j+1}^{k+1}+u_{j+1}^{k}+u_{j+1}^{k-1}),$$

$$u_{j}^{k}\approx\frac{1}{3}(u_{j}^{k+1}+u_{j}^{k}+u_{j}^{k-1}),$$

$$u_{j-1}^{k}\approx\frac{1}{3}(u_{j-1}^{k+1}+u_{j-1}^{k}+u_{j-1}^{k-1});$$

为了避免计算 u 在 $(j+1/2,k)$ 和 $(j-1/2,k)$ 处的值，令

$$a(u_{j+1/2}^{k})\approx a\left(\frac{u_{j+1}^{k}+u_{j}^{k}}{2}\right)=a_{1},$$

$$a(u_{j-1/2}^{k})\approx a\left(\frac{u_{j}^{k}+u_{j-1}^{k}}{2}\right)=a_{2}.$$

于是(1.62)式改写为

$$b(u_{j}^{k})(u_{j}^{k+1}-u_{j}^{k-1})=\frac{2}{3}\frac{\tau}{h^{2}}a_{1}\cdot[(u_{j+1}^{k+1}-u_{j}^{k+1})+(u_{j+1}^{k}-u_{j}^{k})$$
$$+(u_{j+1}^{k-1}-u_{j}^{k-1})]-\frac{2}{3}\frac{\tau}{h^{2}}a_{2}\cdot[(u_{j}^{k+1}-u_{j-1}^{k+1})+(u_{j}^{k}-u_{j-1}^{k})$$
$$+(u_{j}^{k-1}-u_{j-1}^{k-1})]. \tag{1.63}$$

Lees 三层差分格式(1.63)是关于 u_{j}^{k+1} 的线性差分方程组，这样就可利用 $k-1$ 和 k 层上的已知近似值求出 $k+1$ 层上的近似值，避免了解非线性差分方程组。Lees证明了三层格式(1.63)是一个具有截断误差为$O(\tau^{2}+h^{2})$的绝对稳定格式。

8.3 预测校正格式

预测校正方法用于求解非线性抛物型方程一直受到重视，使用该方法可用解线性差分方程代替解非线性差分方程，也能获得较好的数值解精度。

考虑拟线性抛物型方程

$$\frac{\partial^{2}u}{\partial x^{2}}=\psi\left(x,t,u,\frac{\partial u}{\partial x},\frac{\partial u}{\partial t}\right), \tag{1.64}$$

其中，$\partial\psi\Big/\partial\left(\frac{\partial u}{\partial t}\right)\geqslant a_{0}>0$。特别地，这里 ψ 取下述两种形式：

$$\psi = f_1(x,t,u)\frac{\partial u}{\partial t} + f_2(x,t,u)\frac{\partial u}{\partial x} + f_3(x,t,u), \qquad (1.65)$$

$$\psi = g_1\left(x,t,u,\frac{\partial u}{\partial x}\right)\frac{\partial u}{\partial t} + g_2\left(x,t,u,\frac{\partial u}{\partial x}\right). \qquad (1.66)$$

例如，粘性流体力学中的 Burger 方程

$$\frac{\partial u}{\partial t} + u\frac{\partial u}{\partial x} = \nu\frac{\partial^2 u}{\partial x^2}, \qquad x \in (-\infty, +\infty),\ t>0,$$

就属于(1.64)、(1.65)类型；而非线性扩散方程

$$b(u)\frac{\partial u}{\partial t} = \frac{\partial}{\partial x}\left(a(u)\frac{\partial u}{\partial x}\right), \qquad x \in (-\infty, +\infty),\ t>0,$$

就属于(1.64)、(1.66)类型。

对于 ψ 为(1.65)式所示情形，非线性抛物型方程(1.64)的预测校正格式为

$$\begin{cases} \frac{1}{h^2}\delta_x^2 u_j^{k+1/2} \\ \quad = \psi\left[x_j, t_{k+1/2}, u_j^k, \frac{1}{2h}\bar{\delta}_x u_j^k, \frac{2}{\tau}(u_j^{k+1/2} - u_j^k)\right], & (1.67) \\ \frac{1}{2h^2}\delta_x^2(u_j^k + u_j^{k+1}) \\ \quad = \psi\left[x_j, t_{k+1/2}, u_j^{k+1/2}, \frac{1}{4h}\bar{\delta}_x(u_j^k + u_j^{k+1}),\right. \\ \qquad \left.\frac{1}{\tau}(u_j^{k+1} - u_j^k)\right], & (1.68) \end{cases}$$

其中，$\bar{\delta}_x u_j^k = u_{j+1/2}^k - u_{j-1/2}^k$。

对于 ψ 为(1.66)式所示情形，非线性抛物型方程(1.64)的预测校正格式为

预测差分格式仍用(1.67)式，校正差分格式采用

$$\frac{1}{2h^2}\delta_x^2(u_j^k + u_j^{k+1})$$

$$= \psi\left[x_j, t_{k+1/2}, u_j^{k+1/2}, \frac{1}{2h}\delta_x u_j^{k+1/2}, \frac{1}{\tau}(u_j^{k+1} - u_j^k)\right]. \quad (1.69)$$

总之，这两种预测校正格式都是实用的。Douglas 和Jones (1963年)[8]已证明(1.67)和(1.68)式逼近(1.64)和(1.65)式，其截断误差为$O(\tau^2+h^2)$；用(1.67)和(1.69)式的逼近效果是$O(\tau^{3/2}+h^2)$；这两种预测校正格式都是绝对稳定的格式。预测校正格式的计算工作量并不大，因为无论是预测格式还是校正格式都是三对角线性差分方程，在给定边界条件后都可以用追赶法求解。

本章虽然就简单的模型问题介绍了数值求解抛物型方程的种种差分格式，但本章挑选介绍的差分格式就其构思和方法而言是有启发性的。读者学习本章内容时，不仅可将各种差分格式方便有效地推广应用到多个空间自变量的情形中去，而且可以讨论和研究更实用的差分格式。值得指出的是，若在本篇介绍的抛物型方程中固定时间$t=t_k$，则它就成为椭圆型方程。这表明，椭圆型方程的离散化方法，例如有限元方法(见第二篇)，可被有效地引进到抛物型方程的离散化处理中；同样，第三篇中讨论的解离散椭圆方程的高效率高精度算法也可被有效地引进到求解离散抛物型方程中。由于篇幅的限制，本书不能详细地讨论这些问题，读者可根据需要自行拓广应用。

第二章 解双曲型方程的差分方法

大家知道，二阶线性偏微分方程

$$a_{11}\frac{\partial^2 u}{\partial x^2}+2a_{12}\frac{\partial^2 u}{\partial x\partial y}+a_{22}\frac{\partial^2 u}{\partial y^2}+b_1\frac{\partial u}{\partial x}+b_2\frac{\partial u}{\partial y}+cu=f, \tag{2.1}$$

可根据**特征方程**

$$a_{11}dy^2-2a_{12}dxdy+a_{22}dx^2=0$$

分为三大类：若在(x_0, y_0)近旁，$\Delta=a_{12}^2-a_{11}a_{22}<0$，则称方程(2.1)在该点处是**椭圆型**的；若在该点近旁$\Delta=0$，则称方程(2.1)在该点处是**抛物型**的；若在该点处$\Delta>0$，则称方程(2.1)在该点处是**双曲型**的。若方程(2.1)在其定解区域内各点处都是双曲型的，则称方程(2.1)是**严格双曲圆**的。这三类方程描述了不同类型的物理过程。椭圆型方程描述了物理平衡和定常态；抛物型方程描述了诸如物质扩散等不可逆现象；双曲型方程则描述了诸如波的传播等对时间可逆的物理过程。

高阶线性双曲型方程都可写成一阶双曲型方程组

$$\partial U/\partial t+A\partial U/\partial x=F \tag{2.2}$$

的形式，其中$A=(a_{ij})_{n\times n}$，F和U都是$n\times 1$列向量。若A与U无关，F线性地依赖于U，则称(2.2)式为**线性方程组**；若A与U无关，F非线性地依赖于U，则称(2.2)式为**半线性方程型**；若A与U有关，则称(2.2)式是**拟线性方程组**。由方程的分类可知，如果在某点(x, t, U)处，矩阵A有n个互异的实特征值，则称(2.2)式在该点处是双曲型的。本章中，总假定(2.2)式是(严格)双曲型的，即假定方程组(2.2)在定解区域内各点处都是双曲型的。

双曲型方程的初值问题和抛物型方程的初值问题一样，都可转化为常微分方程组的初值问题，这就是说，这两类方程在解的

结构形式和解的稳定性研究方面都有共同之处。另一方面，用差商代替微商所得到的双曲型差分方程和抛物型差分方程在形式上无多大区别，不仅如此，而且这两类差分方程的解结构也是相同的。因此，本篇第一章中关于抛物型方程稳定性定义、研究方法和判别稳定性的有关结论都适合于双曲型差分方程。

本章重点讨论一阶常系数(变系数)双曲型方程(组)的差分方法，并对具有理论和实用价值的拟线性双曲型方程组给予了特别重视。按差分方法的不同特点，本章介绍求解上述方程(组)的直接差分化方法、特征线方法和特征差分格式等，并对求广义解的差分方法作了讨论。

值得指出的是，双曲型方程的**特征**(即**特征线**)在这类方程的定解讨论及其差分方法中都起着重要的作用，读者在学习本章时应注意这一特点。

§1 一阶线性常系数双曲型方程的差分方法

考虑对流方程

$$\partial u/\partial t+a\partial u/\partial x=0,\quad x\in(-\infty,\ +\infty),\quad 0<t\leqslant T \tag{2.3}$$

的初值问题，假定 a 是常数且初始条件为

$$u(x,0)=\varphi(x). \tag{2.4}$$

虽然方程 (2.3) 是简单的，但其差分格式的构造和稳定性讨论是复杂双曲型方程(组)差分方法的基础。

我们知道，方程(2.3)的特征线方向是 $dx/dt=a$，从而可求出其特征线为

$$x-at=\xi,$$

其中 ξ 是常数。在特征线上，(2.3) 式左端正好是 $\dfrac{\partial u}{\partial t}+\dfrac{dx}{dt}\cdot\dfrac{\partial u}{\partial x}=\dfrac{d}{dt}u(x,t)$，故(2.3)式可改写为 $\dfrac{d}{dt}u(x,t)=0$，从

而求出通解为$u(at+\xi,t)=c$。令$t=0$，确定常数c，有$u(\xi,0)=c=\varphi(\xi)=\varphi(x-at)$。这就是说，在特征线上函数值处处相等，解是沿着特征线传播的，即

$$u(x,t)=u(\xi,0)=\varphi(\xi)=\varphi(x-at). \tag{2.5}$$

因此，(2.3)式解的光滑性依赖于初始条件函数的光滑性，当$\varphi(x)\in C^1(R)$时有$u(x,t)\in C^1(R\times T)$。

1.1 古典显格式

不难构造出截断误差为$O(\tau+h)$的古典显格式

$$\frac{1}{\tau}(u_j^{k+1}-u_j^k)+\frac{a}{h}(u_{j+1}^k-u_j^k)=0, \tag{2.6a}$$

和

$$\frac{1}{\tau}(u_j^{k+1}-u_j^k)+\frac{a}{h}(u_j^k-u_{j-1}^k)=0. \tag{2.6b}$$

现用Fourier方法讨论这两个差分格式的稳定性。为此，设$u_j^k=v^k(l)e^{i\sigma x}$并将其代入(2.6a)式，就有

$$v^{k+1}(l)=[1+ar(1-e^{i\sigma h})]v^k(l),\quad r=\frac{\tau}{h},$$

$$\begin{aligned}|G(\sigma,\tau)|^2&=(1+ar(1-\cos\sigma h))^2+a^2r^2\sin^2\sigma h\\&=1+4ar(1+ar)\sin^2\frac{\sigma h}{2}.\end{aligned}$$

此时，Von-Neumann条件是差分格式(2.6a)稳定的充要条件，由上式可知

$$当\left.\begin{cases}a>0\\a<0\ 且\ |ar|\leqslant 1\end{cases}\right\}时，差分格式(2.6a)\begin{cases}不稳定;\\稳定,\end{cases}$$

同样有

$$当\left.\begin{cases}a>0\ 且\ |ar|\leqslant 1\\a<0\end{cases}\right\}时，差分格式(2.6b)\begin{cases}稳定;\\不稳定.\end{cases}$$

若注意到$dx/dt=a$是特征线方向这一事实，则读者立刻可

知古典显格式的稳定性是与特征线方向有关的。这是为什么呢？Courant-Friedrichs-Lewy 结合双曲型方程的依存域，从几何角度分析了差分格式的稳定性和差分解的收敛性。下面以(2.6a)式为例说明之。

由差分格式(2.6a)可知，u_j^{k+1} 依赖于 u_j^k 和 u_{j+1}^k，类推下去，u_j^{k+1} 的值依赖于 $u_j^0, u_{j+1}^0, \cdots, u_{j+k+1}^0$ 的值，如图2.1所示，区间[A, B]称为差分解 u_j^{k+1} 的依赖区域。另外，由(2.5)式知，微分方程解 $u(x_j, t_{k+1})$ 的依赖区域是点 ξ。当 $a>0$ 时，ξ 点落于 A 点左边，差分解依赖区域[A，B]不包含 ξ 点，此时若改变初值 $\varphi(\xi)$ 而不改变[A，B]上的初始值，则由(2.5)式知微分方程的解 $u(x_j, t_{k+1})$要发生变化，但(2.6a)式的差分解u_j^{k+1}却不会发生改变，所见显见，此时 u_j^{k+1}是不会收敛到$u(x_j, t_{k+1})$的。同样可以说明，若 $a<0$，$\xi\in[A, B]$ 时，有 $|a|=\left|\frac{dx}{dt}\right|\leqslant\frac{h}{\tau}$，从而有 $\left|\frac{a\tau}{h}\right|=|ar|\leqslant1$，这正好是(2.6a)式的稳定性条件，此时(2.6a)差分格式相容且稳定，由 Lax 等价定理(参见第一章§3.3)即知，古典显格式(2.6a)的解 u_j^{k+1} 收敛到微分方程 (2.3)式的解 $u(x_j, t_{k+1})$。

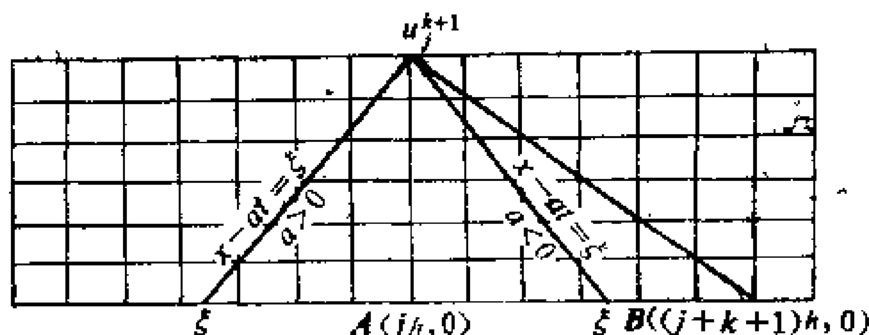


图 2.1

因此，差分解的依赖域包含微分方程解的依赖域是差分解收敛及差分格式稳定的必要条件，这个条件称为 CFL 条件，也可简

称为 Courant 条件。当然不符合 CFL 条件的差分解是不收敛的，其相应的差分格式也是不稳定的。CFL 条件虽然不是差分解收敛(稳定)的充分条件，但它给人们指出了设计差分格式的基本注意事项。另外，读者应该看到，若差分格式满足稳定的充分条件，则可推知该格式一定满足 CFL 条件。

1.2 L-F(Lax-Friedrichs)格式

1954年 Lax 和 Friedrichs 提出 L-F 格式

$$\frac{1}{\tau}\left[u_j^{n+1}-\frac{1}{2}(u_{j-1}^n+u_{j+1}^n)\right]+\frac{a}{2h}[u_{j+1}^n-u_{j-1}^n]=0,$$

其截断误差是 $O(\tau+h^2)$，其传播因子是

$$G(\sigma,\tau)=\cos\sigma h-iar\sin\sigma h,\quad r=\frac{\tau}{h}.$$

因

$$|G|_2^2=\cos^2\sigma h+a^2r^2\sin^2\sigma h=1-(1-a^2r^2)\sin^2\sigma h,$$

所以，由 Von-Neumann 条件得知，当 $|ar|\leqslant1$ 时，L-F 格式稳定。

L-F 格式不象古典显格式那样要考虑特征线走向，因此在一定程度上它还是方便实用的。

1.3 L-W(Lax-Wendroff)格式

1960年 Lax 和 Wendroff 提出一个截断误差为 $O(\tau^2+h^2)$ 的 L-W 格式：

$$u_j^{n+1}=u_j^n-\frac{ar}{2}(u_{j+1}^n-u_{j-1}^n)+\frac{a^2r^2}{2}(u_{j-1}^n-2u_j^n+u_{j+1}^n).$$

在解充分光滑时，由于

$$\frac{\partial^2u}{\partial t^2}=\frac{\partial}{\partial t}\left(\frac{\partial u}{\partial t}\right)=\frac{\partial}{\partial t}\left(-a\frac{\partial u}{\partial x}\right)$$

$$=-a\frac{\partial}{\partial x}\left(\frac{\partial u}{\partial t}\right)=a^2\frac{\partial^2u}{\partial x^2},$$

并且

$$u(x_j, t_{k+1}) = u(x_j, t_k) + \tau\left[\frac{\partial u}{\partial t}\right]_j^k + \frac{\tau^2}{2}\left[\frac{\partial^2 u}{\partial t^2}\right]_j^k + O(\tau^3)$$

$$= u(x_j, t_k) - a\tau\left[\frac{\partial u}{\partial x}\right]_j^k + \frac{a^2\tau^2}{2}\left[\frac{\partial^2 u}{\partial x^2}\right]_j^k + O(\tau^3),$$

从而可知 L-W 格式的截断误差是 $O(\tau^2 + h^2)$。

不难求出 L-W 格式的传播因子是

$$G(\sigma,\tau) = 1 - 2a^2r^2\sin^2\frac{\sigma h}{2} - iar\sin\sigma h,$$

所以有

$$|G(\sigma,\tau)|^2 = \left(1 - 2a^2r^2\sin^2\frac{\sigma h}{2}\right)^2 + a^2r^2\sin^2\sigma h$$

$$= \left(1 - 2a^2r^2\sin^2\frac{\sigma h}{2}\right)^2 + 4a^2r^2\sin^2\frac{\sigma h}{2}\cos^2\frac{\sigma h}{2}$$

$$= 1 - 4a^2r^2(1 - a^2r^2)\sin^4\frac{\sigma h}{2}.$$

故知当 $|ar| \leqslant 1$ 时 L-W 格式是稳定的。

L-W 格式和 L-F 格式一样，都是二层显格式，都无需顾及特征线的走向。

1.4 跳蛙(Leap-frog)格式

跳蛙格式

$$\frac{1}{2\tau}[u_j^{k+1} - u_j^{k-1}] + \frac{a}{2h}[u_{j+1}^k - u_{j-1}^k] = 0$$

是一个截断误差为 $O(\tau^2 + h^2)$ 的三层格式。

为了分析其稳定性，令 $v_j^k = u_j^{k-1}$，上式就可化为一个二层差分方程组：

$$\begin{pmatrix} u \\ v \end{pmatrix}_j^{k+1} = \begin{pmatrix} -ar & 0 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} u \\ v \end{pmatrix}_{j+1}^k + \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} u \\ v \end{pmatrix}_j^k + \begin{pmatrix} ar & 0 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} u \\ v \end{pmatrix}_{j-1}^k,$$

再用 Fourier 分析方法，就可得到传播矩阵

$$G(\sigma,\tau) = \begin{pmatrix} -i2ar\sin\sigma h & 1 \\ 1 & 0 \end{pmatrix}.$$

G的特征值为

$$\mu = -iar\sin\sigma h \pm \sqrt{1 - a^2r^2\sin^2\sigma h}.$$

当$|ar|>1$时，对于满足$a^2r^2\sin^2\sigma h>1$的σh，G的两个特征值都是纯虚数，又因为两根之和的模$|i2ar\sin\sigma h|>2$，所以由本篇第一章中引理1.1可知，G的两个特征值中必有一个的绝对值大于1．也就是说，当$|ar|>1$时，跳蛙格式不稳定。

当$|ar|\leqslant 1$且$\sigma h \neq n\pi$时，传播矩阵G有两个互异的特征值，利用定理1.7中的1°可知，此时跳蛙格式是稳定的。

当$|ar|\leqslant 1$且$\sigma h = n\pi$时，虽然G的特征值有重根，但此时传播矩阵为

$$G(n\pi,\tau) = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix},$$

且$\{G^k(n\pi,\tau)\}$是一致有界的．利用稳定性的定义，即知跳蛙格式在这种情况下也是稳定的。

总之，跳蛙格式的稳定性条件是$|ar|\leqslant 1$。

1.5 隐格式

普通的隐格式

$$\frac{1}{\tau}[u_j^k - u_j^{k-1}] + \frac{a}{2h}[u_{j+1}^k - u_{j-1}^k] = 0$$

的截断误差为$O(\tau+h^2)$，该格式绝对稳定。

为了提高差分格式的精度，现将上式修改为

$$\frac{1}{\tau}[u_j^k - u_j^{k-1}] + \frac{1}{2}\cdot\frac{a}{2h}[(u_{j+1}^k - u_{j-1}^k) + (u_{j+1}^{k-1} - u_{j-1}^{k-1})] = 0.$$

此格式有时也称作 **Cranb-Nicholson** 格式。为了估计其截断误差，现将它的各项在$(x_j, t_{k-1/2})$处作 Taylor 展开：

$$\frac{1}{\tau}[u(x_j, t_k) - u(x_j, t_{k-1})] = \left(\frac{\partial u}{\partial t}\right)_j^{k-\frac{1}{2}} + O(\tau^2),$$

$$\frac{1}{2h}[u(x_{j+1}, t_k) - u(x_{j-1}, t_k)] = \left(\frac{\partial u}{\partial x}\right)^{k-\frac{1}{2}} + \frac{\tau}{2}\left(\frac{\partial^2 u}{\partial x \partial t}\right)_j^{k-\frac{1}{2}} + O(\tau^2 + h^2),$$

$$\frac{1}{2h}[u(x_{j+1}, t_{k-1}) - u(x_{j-1}, t_{k-1})] = \left(\frac{\partial u}{\partial x}\right)_j^{k-\frac{1}{2}} - \frac{\tau}{2}\left(\frac{\partial^2 u}{\partial x \partial t}\right)_j^{k-\frac{1}{2}} + O(\tau^2 + h^2),$$

由此不难知道，该格式的截断误差是 $O(\tau^2 + h^2)$。

稳定性讨论也是不难的，先求出传播因子

$$G(\sigma, \tau) = \frac{1 - \frac{ar}{4}(e^{i\sigma h} - e^{-i\sigma h})}{1 + \frac{ar}{4}(e^{i\sigma h} - e^{-i\sigma h})} = \frac{1 - i\frac{ar}{2}\sin\sigma h}{1 + i\frac{ar}{2}\sin\sigma h},$$

并有$|G|^2 \leqslant 1$，所以该格式是绝对稳定的。

1.6 特征线方法

由 (2.5) 式得到启发，若不用矩形网格而用斜交网格，即用由斜率为 $dx/dt = a$ 的平行直线族和 $t_k = k\tau$ 的平行直线族组成的斜交网格，则利用(2.5)式立即有 $u_j^k = u(x_j, t_k) = \varphi(x_j - ak\tau)$。

这是准确求解的办法，参见图2.2.

1.7 特征差分格式

利用矩形网格，假定 CFL 条件成立，并考虑到方程(2.3)的解在特征线上为常数这一事实，现利用特征和插值来构造差分格式.

不妨设 $a>0$，见图 2.3. 一方面，$u(P)=u(Q)$，Q 不是网格节点；另一方面，$u(Q)$ 利用 $u(B)$ 和 $u(C)$ 作线性插值得到，这样有

$$u(P)=u(Q)=\left(1-a\frac{\tau}{h}\right)u(C)+a\frac{\tau}{h}u(B);$$

由此导出的差分格式

$$u_j^{k+1}=u_j^k-ar(u_j^k-u_{j-1}^k)$$

与古典显格式(2.6b)同.

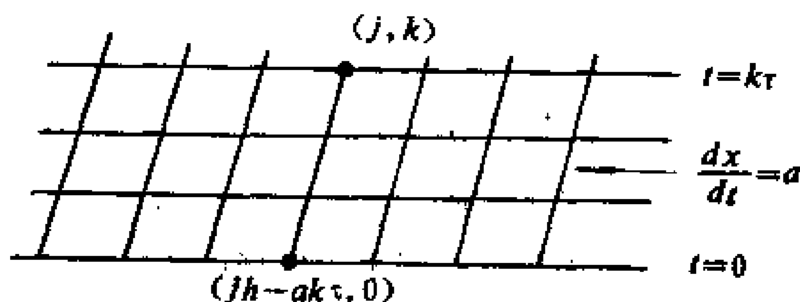


图 2.2

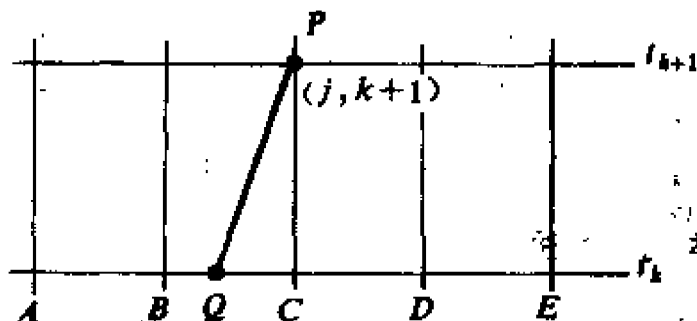


图 2.3

如果改用 B、D 两点作线性插值表示 $u(Q)$，则可导出 Lax

-Friedrichs 格式。

如果采用 B、C、D 三点作二次插值表示 $u(Q)$，有

$$u(P)=u(Q)=u(C)-ar[u(C)-u(B)]$$

$$-\frac{ar}{2}(1-ar)[u(B)-2u(C)+u(D)],$$

则可导出 Lax-Wandroff 格式。

§2　一阶线性常系数双曲型方程组的差分方法

考虑常系数双曲型方程组

$$\partial U/\partial t+A\partial U/\partial x=0 \tag{2.7}$$

的初值问题，初始条件取 $U(x,0)=U_0(x)$，其中 $U=U(x,t)=(u_1(x,t),\cdots,u_n(x,t))^T$，$A$ 是 $n\times n$ 阶常数矩阵。因为方程组(2.7)是双曲型的，A 有 n 个不同的实特征值，那么存在非奇异常数矩阵 S，使有

$$SAS^{-1}=\Lambda \text{或} A=S^{-1}\Lambda S,\quad \Lambda=\begin{pmatrix}\lambda_1 & & 0\\ & \ddots & \\ 0 & & \lambda_n\end{pmatrix}.$$

若令 $W=SU$，则方程组(2.7)的初值问题可化为

$$\partial W/\partial t+\Lambda\partial W/\partial x=0 \tag{2.8}$$

的初值问题，这里初值可写为 $W(x,0)=W_0(x)$。注意，方程组(2.8)是由 n 个彼此无关的方程构成，每个方程都与(2.3)式同形。因此，利用(2.3)和(2.4)式所描述的对流方程的适定性，就可确定常系数双曲型方程组的适定性：若 $U_0(x)\in C^1(R^n)$，则方程组(2.7)的初值问题有唯一解，且 $U(x,t)\in C^1(R^n\times T)$。

由以上讨论可知，§1中关于对流方程的各种差分格式都可以推广到一阶线性常系数方程组的情形中。本节仅选择§1中几个有典型实用意义的差分格式进行推广，关于方程组(2.7)的初始条件和边值条件的处理将在§4中讨论。

2.1 L-F格式

将一个方程的Lax-Friedrichs格式直接推广到方程组情形，就有

$$\frac{1}{\tau}\left[U_j^{k+1}-\frac{1}{2}(U_{j+1}^k+U_{j-1}^k)\right]$$

$$+A\frac{1}{2h}[U_{j+1}^k-U_{j-1}^k]=0,$$

这仍是一个精度为 $O(\tau+h^2)$ 的差分格式。关于稳定性讨论也是方便的，只要令$U_j^k=V^k e^{i\sigma x}$，$r=\tau/h$，代入上式即可求得传播矩阵

$$G(\sigma,\tau)=\frac{1}{2}(e^{i\sigma h}+e^{-i\sigma h})I$$

$$+\frac{r}{2}(e^{-i\sigma h}-e^{i\sigma h})A$$

$$=\cos\sigma h\cdot I-ir\sin\sigma h\cdot A,$$

这是一个关于矩阵A的多项式。因此，若设 $\lambda_l(l=1,\cdots,n)$ 是A的特征值，则可顺利写出G的特征值$\mu_l(G)$及模：

$$\mu_l(G)=\cos\sigma h-ir\sin\sigma h\cdot\lambda_l,$$

$$|\mu_l|^2=\cos^2\sigma h+r^2\lambda_l^2\sin^2\sigma h$$

$$=1-(1-r^2\lambda_l^2)\sin^2\sigma h,\quad l=1,\cdots,n.$$

显然，如果$r\rho(A)\leqslant 1$，则L-F格式适合Von-Neumann条件。但因方程组(2.7)是双曲型的，A有n个互异的实特征值，所以由定理1.7中的1°即知，当$r\rho(A)\leqslant 1$时L-F格式稳定。

2.2 L-W格式

我们仍把单个方程的L-W格式推广到方程组情形，即有

$$U_j^{k+1}=U_j^k-\frac{r}{2}A(U_{j+1}^k-U_{j-1}^k)$$

$$+\frac{r^2}{2}A^2(U_{j-1}^n-2U_j^n+U_{j+1}^n),$$

这里$r=\tau/h$。容易看出 L-W 格式是二阶精度的格式。其传播矩阵是

$$G(\sigma,\tau)=I-ir\sin\sigma h\cdot A-r^2(1-\cos\sigma h)A^2.$$

记λ_l和μ_l分别为A和G的特征值，则有

$$\mu_l(G)=1-ir\lambda_l\sin\sigma h-r^2\lambda_l^2(1-\cos\sigma h),$$

$$|\mu_l|^2=1-4r^2\lambda_l^2(1-r^2\lambda_l^2)\sin^4\frac{\sigma h}{2},\qquad l=1,\cdots,n.$$

此时，von-Neumann 条件仍是差分格式稳定的充分条件，故当$r\rho(A)\leqslant 1$时 L-W 格式稳定。

2.3 特征差分格式

由§1.7看出，一个方程的特征差分格式(采用线性插值情形)同于古典显格式(2.6a)或(2.6b)，但将此格式直接推广到方程组(2.7)有困难，主要是由于A的特征值可正可负，我们无法判断特征线的走向的缘故。

为此，先讨论方程组(2.8)的特征差分格式。注意方程组(2.8)已经不是耦合形式了，它相当于n个独立方程

$$\partial w^{(l)}/\partial t+\lambda_l\partial w^{(l)}/\partial x=0,\qquad l=1,\cdots,n,$$

其中$w^{(l)}$是W的第l个分量，λ_l是A的第l行对角元素，这个方程可以按对流方程的特征差分格式(如(2.6a)和(2.6b)式)处理。为书写方便，省去$w^{(l)}$的上标l，则上式的特征差分格式是

$$\frac{1}{\tau}[w_j^{n+1}-w_j^n]+\frac{\lambda_l}{h}[w_j^n-w_{j-1}^n]=0,\qquad 当\ \lambda_l\geqslant 0时,$$

$$\frac{1}{\tau}[w_j^{n+1}-w_j^n]+\frac{\lambda_l}{h}[w_{j+1}^n-w_j^n]=0,\qquad 当\ \lambda_l<0.$$

把上述两种情况写成统一形式

$$\frac{1}{\tau}[w_j^{n+1}-w_j^n]+\frac{\lambda_l}{2h}[w_{j+1}^n-w_{j-1}^n]$$

$$=\frac{|\lambda_l|}{2h}[w_{j+1}^n-2w_j^n+w_{j-1}^n].$$

再把它写成向量形式的特征差分格式：

$$W_j^{n+1}=W_j^n-\frac{r}{2}\Lambda(W_{j+1}^n-W_{j-1}^n)$$
$$+\frac{r}{2}|\Lambda|(W_{j+1}^n-2W_j^n+W_{j-1}^n). \qquad (2.9)$$

这个特征差分格式自动适应了特征值符号，从而自动适应了 CFL 条件。这样就可以把 (2.9) 式推广为方程组 (2.7) 的特征差分格式：

$$U_j^{n+1}=U_j^n-\frac{r}{2}A(U_{j+1}^n-U_{j-1}^n)$$
$$+\frac{r}{2}|A|(U_{j+1}^n-2U_j^n+U_{j-1}^n), \qquad (2.10)$$

其中 $r=\tau/h$，$|A|=S|\Lambda|S^{-1}$.

下面先证特征差分格式(2.9)的稳定性。(2.9)式的传播矩阵是

$$G_1(\sigma,\tau)=I-ir\sin\sigma h\Lambda+r(\cos\sigma h-1)\cdot|\Lambda|,$$

其特征值

$$\mu_l(G_1)=1-ir\lambda_l\sin\sigma h+r|\lambda_l|(\cos\sigma h-1),$$
$$|\mu_l|^2=(1-2r|\lambda_l|\sin^2(\sigma h/2))^2+r^2\lambda_l^2\sin^2\sigma h$$
$$=1-4r|\lambda_l|(1-r|\lambda_l|)\sin^2(wh/2),\ l=1,\cdots,n.$$

因 Λ 是对角阵，所以 Von-Neumann 条件是差分格式(2.9)稳定的充分条件。故可确定，当 $r\rho(\Lambda)\leqslant 1$时，有 $\rho(G_1)\leqslant 1$，差分格式(2.9)稳定。此时$\|G_1^n(\sigma,\tau)\|\leqslant M_1$是显然的。

再证特征差分格式(2.10)的稳定性。设 (2.10) 式的传播矩阵为$G_2(\sigma,\tau)$。因$A=S\Lambda S^{-1}$，所以有$G_2=SG_1S^{-1}$，$G_2^n=SG_1^nS^{-1}$，可推知$\|G_2^n\|\leqslant M_2$。故当 $r\rho(A)\leqslant 1$时，特征差分格式 (2.10) 也

是稳定的。

§3 一阶变系数双曲型方程(组)的差分方法

一阶变系数双曲型方程

$$b(x,t)\frac{\partial u}{\partial t}+a(x,t)\ \frac{\partial u}{\partial x}=c(x,t),\qquad |x|<\infty,\ 0<t\leqslant T \tag{2.11}$$

的特征线方向是

$$\frac{dx}{dt}=\frac{a(x,t)}{b(x,t)}\text{或}\begin{vmatrix} a(x,t) & b(x,t)\\ dx & dt\end{vmatrix}=0, \tag{2.12}$$

这表明特征线是一曲线；解沿特征线微分关系是

$$\frac{du}{dt}=\frac{c(x,t)}{b(x,t)}\text{或}\begin{vmatrix} c(x,t) & b(x,t)\\ du & dt\end{vmatrix}=0, \tag{2.13}$$

这表明，沿特征线方向，方程(2.11)可化为常微分方程。因此只要初始值曲线不与特征线平行，方程(2.11)的初值问题的适定性是容易确定的。

对于变系数双曲型方程组：

$$B(x,t)\partial U/\partial t+A(x,t)\partial U/\partial x=C(x,t), \tag{2.14}$$

其中，$U(x,t)=(u_1(x,t),u_2(x,t),\cdots,u_n(x,t))^T$，$C(x,t)=(c_1(x,t),c_2(x,t),\cdots,c_n(x,t))^T$，$A(x,t)$和$B(x,t)$都是与$x$和$t$有关的$n\times n$阶矩阵。人们可以从两个方面研究方程组(2.14)。

一方面是，研究解沿着特征线传播的本质(见§5.1)，就可确定(2.14)式所决定的特征线方向场和解沿特征线方向的微分关系，这些结果类似于一个方程相应情形的(2.12)和(2.13)式。

另一方面是，利用(2.14)式是严格双曲型方程组这一本质，对其作相似变换(仿照一阶线性常系数双曲型方程组情形)，那么(2.14)式一定可改写为

$$\partial W/\partial t+\Lambda(x,t)\partial W/\partial x=F \tag{2.15}$$

的形式，其中W和F都是关于x和t的n阶列向量，$A(x,t)$是对角阵。$A(x,t)$的元素正好反映了n条特征曲线的方向场。

一阶变系数双曲型方程组确定了n条不同曲线的特征线，这是与常系数双曲型方程组的情形不相同的。

现以$n=2$的情形为例，说明方程组特征线把定解区域分为依线赖、决定域和影响域。如图2.4所示，定解区域中，点$M(x,t)$处有2条曲特征线l_1和l_2通过，显然曲边三角形MAB中各点函数值都是由初值线AB上的数值唯一确定的，故称AB为**依赖域**；称曲边三角形MAB为**决定域**；由l_2'、AB、和l_1'所围区域中各点函数值都受到AB上初值的影响，称此区域是**影响域**。

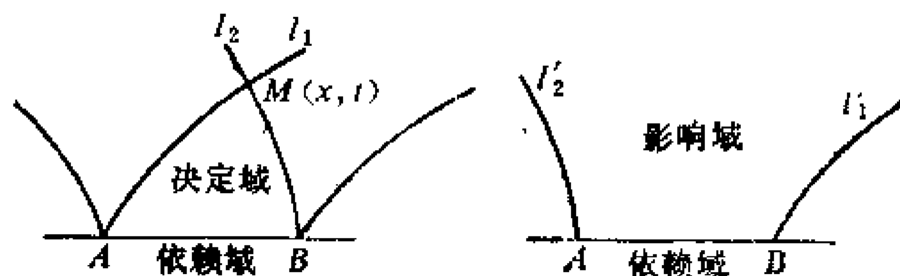


图2.4

常系数双曲型方程(组)的种种差分格式都可以推广到变系数情形，然而稳定性分析将出现复杂性。本节将用几个例子着重指明如何进行推广，以及如何解决稳定性分析，而不准备将前面出现的各种方法面面俱到地罗列。

3.1 稳定性分析的冻结系数方法和能量估计

对于变系数对流方程

$$\partial u/\partial t + a(x,t)\partial u/\partial x = 0, \qquad |x|<\infty, \ 0<t\leqslant T,$$

可仿照常系数情形，直接写出特征差分格式

$$\begin{cases} \dfrac{1}{\tau}(u_j^{n+1}-u_j^n)+\dfrac{a_j^n}{h}(u_j^n-u_{j-1}^n)=0, & a_j^n\geqslant 0, \\ \dfrac{1}{\tau}(u_j^{n+1}-u_j^n)+\dfrac{a_j^n}{h}(u_{j+1}^n-u_j^n)=0, & a_j^n<0, \end{cases} \tag{2.16}$$

为了分析该格式的稳定性，我们先把a_j^k看作常数，于是(2.16)式就是常系数差分方程，用 Fourier 分析法求得稳定性条件是$|a_j^k\tau/h|\leqslant 1$. 然而这个结果只能近似反映节点(x_j, t_k)附近的情形。因此，再对一切的x_j考虑，就有差分方程(2.16)的稳定性条件

$$\frac{\tau}{h}\max_j|a_j^k|\leqslant 1.$$

这就是**冻结系数法**。该方法很简单，而且一般能较好地反应稳定性结果，然而这不是严格的方法。

用**能量估计**的办法分析变系数差分格式的稳定性是普遍适用的，现仍以(2.16)式为例作能量估计。

先把(2.16)式中的第一式改写为

$$u_j^{k+1}=(1-ra_j^k)u_j^k+ra_j^ku_{j-1}^k \quad (r=\tau/h),$$

两边同乘u_j^{k+1}并使用条件$0<\frac{\tau}{h}\max\limits_j a_j^k\leqslant 1$，就有

$$(u_j^{k+1})^2\leqslant\frac{1-ra_j^k}{2}[(u_j^k)^2+(u_j^{k+1})^2]+\frac{ra_j^k}{2}[(u_{j-1}^k)^2+(u_j^{k+1})^2]$$

$$=\frac{1}{2}(u_j^{k+1})^2+\frac{1-ra_j^k}{2}(u_j^k)^2+\frac{ra_j^k}{2}(u_{j-1}^k)^2,$$

$$(u_j^{k+1})^2\leqslant(1-ra_j^k)(u_j^k)^2+ra_j^k(u_{j-1}^k)^2$$

$$=(u_j^k)^2-ra_j^k(u_j^k)^2+ra_j^k(u_{j-1}^k)^2;$$

将上式两边再同乘h，对j求和并令$\|u^k\|^2=\sum\limits_{j=-\infty}^{+\infty}(u_j^k)^2h$，就有

$$\|u^{k+1}\|^2\leqslant\|u^k\|^2+r\sum_{j=-\infty}^{\infty}(a_{j+1}^k-a_j^k)(u_j^k)^2h;$$

再令$c=\sup\limits_{k,j}|(a_{j+1}^k-a_j^k)/h|$，最后得到

$$\|u^{k+1}\|^2\leqslant(1+c\tau)\|u^k\|^2.$$

这个结果表明，(2.16)式中的第一式是按模$\|\cdot\|$稳定的。

用同样的办法可证得(2.16)式的第二式的稳定性。

总之可得出结论：若$a(x,t)$关于x是Lipschitz连续的，且$\frac{\tau}{h}\max_j|a_j^k|\leqslant 1$，则差分格式(2.16)稳定。

3.2 变系数方程的特征线方法

对于变系数方程(2.11)，可确定出特征线方向(2.12)式和沿特征线解的微分关系(2.13)式，原则上，利用这两个关系式能求出方程(2.11)解的表达式。例如，首先利用$dx/dt=a(x,t)/b(x,t)$定出$x=x(t)$，再将这个特征线表达式代入$du/dx=c(x,t)/b(x,t)$并积分，就可求出u在特征上的表达式了。然而对于复杂的变系数情形，用这种办法是很难求出方程解的一般解析表达式的。可是这一思想用于数值方法(即**特征线法**)恰是简单易行，而且可以推广到数值求解更复杂的拟线性双曲型方程情形。

简言之，**特征线法**就是用Euler折线法近似(2.12)式和(2.13)式。所以有差分格式

$$
\begin{cases}(x_j^{k+1}-x_j^k)/(t_{k+1}-t_k)=a(x_j^k,t_k)/b(x_j^k,t_k), & k=0,1,\cdots,\\ (u_j^{k+1}-u_j^k)/(t_{k+1}-t_k)=c(x_j^k,t_k)/b(x_j^k,t_k), & j=0,1,\cdots.\end{cases}
$$

特征线方法的计算过程可参见图2.5，图中AB假设为初值线，实曲线为特征线，虚折线表示用特征线法的计算过程。据

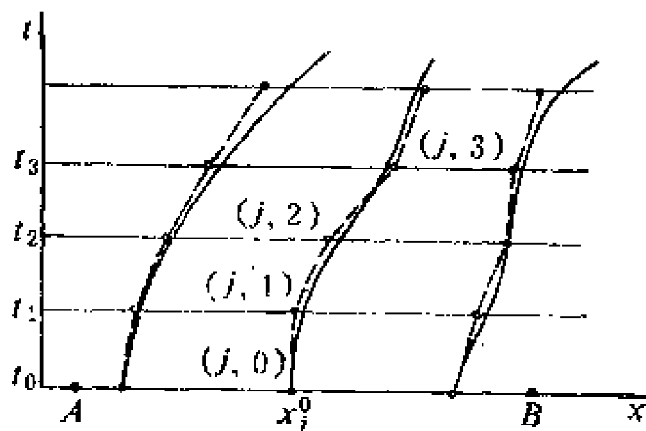


图2.5

Euler 折线法精度分析，可知特征线法的计算精度为 $O(\tau)$。

一阶常系数线性方程和方程组的有关差分格式都可以直接推广到变系数情形，稳定性分析一般采用冻结系数法和能量估计法。此节不再反复列举，读者可自行练习。

虽然本节只是就两个例子介绍了特征差分格式和特征线法，但关于方程组情形的其它差分格式及边界条件的处理方法，读者还可在§4中得到部分补充和启发。至于特征线法，还可在§5中得到更深刻的理解。

§4 二阶双曲型方程的差分方法

最简单最典型的二阶双曲型方程是一维**波动方程**

$$\partial^2 u/\partial t^2 = a^2\partial u^2/\partial x^2, \tag{2.17}$$

其中常数 $a>0$。根据二阶偏微分方程理论，与(2.17)式相应的特征方程为 $dx^2-a^2dt^2=0$，利用特征方向 $dx/dt=\pm a$，即可求得两族特征线：

$$x-at=\xi_1, \qquad x+at=\xi_2.$$

若用 u 沿特征线的偏导数分别表示 $\dfrac{\partial^2 u}{\partial t^2}$ 和 $\dfrac{\partial^2 u}{\partial x^2}$：

$$\begin{aligned}\frac{\partial^2 u}{\partial t^2} &= \frac{\partial}{\partial t}\left(\frac{\partial u}{\partial \xi_1}\frac{\partial \xi_1}{\partial t}+\frac{\partial u}{\partial \xi_2}\frac{\partial \xi_2}{\partial t}\right)\\ &= a^2\left(\frac{\partial^2 u}{\partial \xi_1^2}-2\frac{\partial^2 u}{\partial \xi_1\partial \xi_2}+\frac{\partial^2 u}{\partial \xi_2^2}\right),\end{aligned}$$

$$\begin{aligned}\frac{\partial^2 u}{\partial x^2} &= \frac{\partial}{\partial x}\left(\frac{\partial u}{\partial \xi_1}\frac{\partial \xi_1}{\partial x}+\frac{\partial u}{\partial \xi_2}\frac{\partial \xi_2}{\partial x}\right)\\ &= \frac{\partial^2 u}{\partial \xi_1^2}+2\frac{\partial^2 u}{\partial \xi_1\partial \xi_2}+\frac{\partial^2 u}{\partial \xi_2^2},\end{aligned}$$

把它们代入(2.17)式，则得到 $\partial^2 u/\partial \xi_1\partial \xi_2=0$，从而得到波动方程的通解形式为

$$u = f_1(\xi_1) + f_2(\xi_2) = f_1(x - at) + f_2(x + at).$$

若再考虑初始条件

$$u(x,0) = \varphi_0(x), \quad u_x(x,0) = \varphi_1(x), \qquad |x| < \infty,$$

则可确定函数 f_1 和 f_2，且求出解的表达式

$$u(x,t) = \frac{1}{2}[\varphi_0(x - at) + \varphi_0(x + at)] + \frac{1}{2a}\int_{x-at}^{x+at} \varphi_1(\xi)\,d\xi,$$

这就是大家熟知的**达朗贝尔**(D. Alembert)**公式。**

利用 D. Alembert 公式，很容易理解点 $M(x_0, t_0)$ 的决定域是 $\triangle MAB$，依赖域是$[AB]$(参见图2.6a))；也很容易理解点 x_0 的影响域(参见图2.6b))：

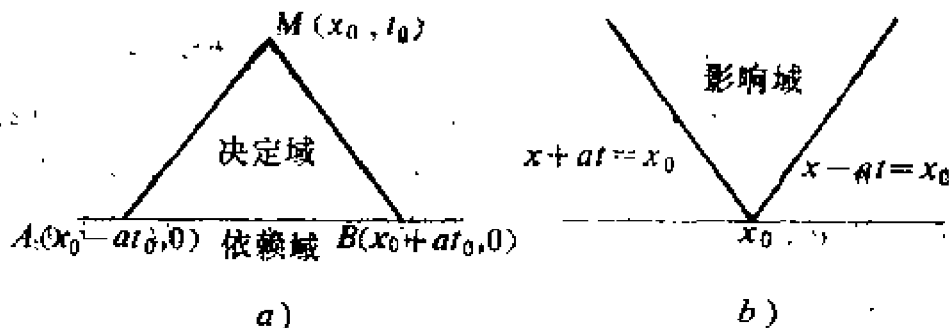


图2.6

正如本章开头所述，高阶方程可化为一阶方程组。这里，只要令 $v = \partial u/\partial t$，$w = a\partial u/\partial x$，$U = (v, w)^T$，波动方程(2.17)就可写成一阶线性方程组

$$\partial U/\partial t + A\partial U/\partial x = 0 \tag{2.18}$$

的形式，其中$A = \begin{pmatrix} 0 & -a \\ -a & 0 \end{pmatrix}$。因$A$有两个互异特征值($\pm a$)，所以也存在非奇异矩阵$S$，使$A$化为对角形$\Lambda$，即

$$S = \frac{1}{2}\begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix}, \quad \Lambda = S^{-1}AS = \begin{pmatrix} a & 0 \\ 0 & a \end{pmatrix}.$$

所以，(2.18)式是严格双曲型方程组。

4.1 二阶方程的显格式

仿抛物型方程的直接差分化方法，可构造二阶方程(2.17)的显格式为

$$\frac{1}{\tau^2}[u_j^{k+1}-2u_j^k+u_j^{k-1}]=\frac{a^2}{h^2}[u_{j-1}^k-2u_j^k+u_{j+1}^k]. \quad (2.19)$$

这是一个截断误差为 $O(\tau^2+h^2)$ 的三层显格式。为了用Fourier方法分析其稳定性，需将其化成两层格式。容易验证(2.19)式与

$$\begin{cases}\frac{1}{\tau}(v_j^{k+1}-v_j^k)=\frac{a}{h}(w_{j+1/2}^k-w_{j-1/2}^k),\\ \frac{1}{\tau}(w_j^{k+1}-w_j^k)=\frac{a}{h}(v_{j+1/2}^{k+1}-v_{j-1/2}^{k+1}),\end{cases} \quad (2.20)$$

是等价的。其中 $v_j^k=\frac{1}{\tau}(u_j^k-u_j^{k-1})$,

$$w_j^k=\frac{a}{h}(u_{j+1/2}^k-u_{j-1/2}^k).$$

现用Fourier分析法分析(2.20)式的稳定性。令 $v_j^k=V_1^k(l)\cdot e^{i\sigma x}$, $w_j^k=V_2^k(l)e^{i\sigma x}$，代入(2.20)式就有

$$\begin{pmatrix}1 & 0\\ -2ci & 1\end{pmatrix}\begin{pmatrix}V_1^{k+1}\\ V_2^{k+1}\end{pmatrix}=\begin{pmatrix}1 & 2ci\\ 0 & 1\end{pmatrix}\begin{pmatrix}V_1^k\\ V_2^k\end{pmatrix},$$

或

$$\begin{pmatrix}V_1^{k+1}\\ V_2^{k+1}\end{pmatrix}=\begin{pmatrix}1 & ic\\ ic & 1-c^2\end{pmatrix}\begin{pmatrix}V_1^k\\ V_2^k\end{pmatrix},$$

其中，$r=\frac{\tau}{h}$, $c=2\mathrm{arcsin}\frac{\sigma h}{2}$.

传播矩阵是

$$G=\begin{pmatrix}1 & ic\\ ic & 1-c^2\end{pmatrix},$$

其特征方程是

$$\mu^2-(2-c^2)\mu+1=0,$$

特征根是

$$\mu_{1,2}=1-2a^2r^2\sin^2\frac{\sigma h}{2}\pm i\sqrt{\left(1-a^2r^2\sin^2\frac{\sigma h}{2}\right)4a^2r^2\sin^2\frac{\sigma h}{2}}.$$

据引理1.1，$|\mu|\leqslant 1$ 的充要条件是 $|2-c^2|\leqslant 2$，故当 $ar\leqslant 1$ 时 $|\mu|\leqslant 1$，符合 von Neumann 条件。但因为传播矩阵 G 不是正规阵，所以还需仔细地分析稳定性条件。

当 $ar<1$ 且 $\sigma h\neq 2n\pi$ 时，G 有两个互异的特征值且 $|\mu_{1,2}|=1$，据定理1.7，差分格式(2.20)此时是稳定的。

当 $ar<1$ 且 $\sigma h=2n\pi$ 时，$|\mu_{1,2}|=1$，传播矩阵 G 是单位阵，直接运用稳定性概念，此时差分格式(2.20)也是稳定的。

当 $ar=1$ 时，G 的特征值 $\mu_1=\mu_2=-1$，它有两个初等因子；特别地取 $\sigma h=\pi$，此时有相似变换 S 存在，使

$$G=S\begin{pmatrix}-1 & 1\\ 0 & -1\end{pmatrix}S^{-1},$$

$$G^k=S\begin{pmatrix}(-1)^k & (-1)^{k-1}k\\ 0 & (-1)^k\end{pmatrix}S^{-1},\quad k=1,2,\cdots,$$

显然 $\{G^k\}$ 无界，所以当 $ar=1$ 时差分格式(2.20)是不稳定的。

总之，差分格式(2.19)或(2.20)式稳定的充要条件是 $ar<1$。再据定理1.3即知，当 $ar<1$ 时显格式(2.19)的差分解收敛于波动方程(2.17)的解。

当 $ar\leqslant 1$ 时，显格式(2.19)符合 CFL 条件。事实上，从差分方程(2.19)看出，u_j^k 依赖前两层的值 u_{j-1}^{k-1}、u_j^{k-1}、u_{j+1}^{k-1} 和 u_j^{k-2}，依此类推，可知 u_j^k 最终依赖于初始层上的下列值：

$$u_{j-k}^0,\ u_{j-k+1}^0,\cdots,u_j^0,\cdots,u_{j+k-1}^0,\ u_{j+k}^0.$$

因此称 x 轴上含于区间 $[x_{j-k},\ x_{j+k}]$ 中的两格点为差分解 u_j^k 的依存域，它是在 x 轴上被通过 (x_j,t_k) 的两条直线

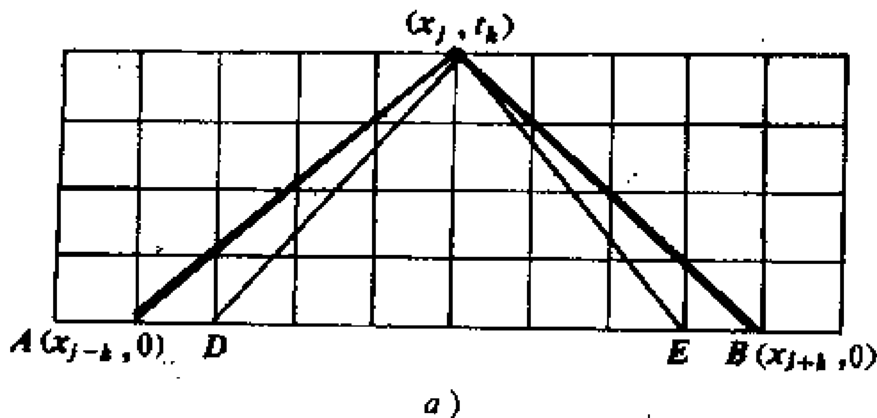
$$x-x_j=\pm\frac{h}{\tau}(t-t_k)$$

切劏下来的区间[A，B]上的网格点。另外，过(x_j, t_k)，微分方程(2.17)的两条特征线是

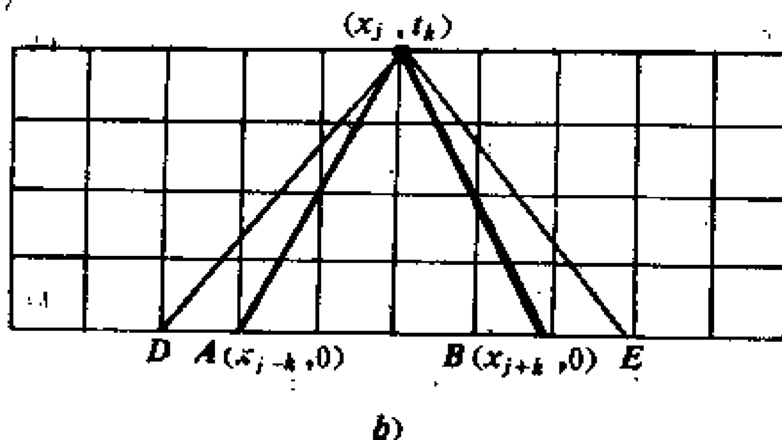
$$x - x_j = \pm a(t - t_k),$$

[D，E]是微分方程解$u(x_j, t_k)$的依存域。显然，当$ar \leqslant 1$时，[A, B]$\supseteq$[D, E]，参见图2.7a)，显格式(2.19)可能稳定。据前面的稳定性分析看出，$ar < 1$时显格式(2.19)是稳定的，但$ar = 1$时显格式(2.19)是不稳定的。

当$r > 1$时，差分解的依存域[A, B]不能包含微分方程解的依存域[D, E]，参见图2.7b)，此时差分解是不稳定的。



a)



b)

图2.7

4.2 定解条件的处理

逼近二阶方程(2.17)的显格式和隐格式还有多种，这里不再

一一列举。在解差分方程时，必须将微分方程的离散方程和边界条件的离散方程联立求解，这就要求这两类离散方程的截断误差相匹配，否则会影响差分解的精度。

对初始条件

$$\begin{cases} u(x,0)=\varphi(x), \\ u_t(x,0)=\psi(x), \quad x\in(-\infty,+\infty) \end{cases}$$

来说，其第一个方程的离散形式

$$u(x_j,0)=u_j^0=\varphi(x_j)$$

可适合各种精度的差分格式，但第二个边界条件的离散方程 $(u_j^1 - u_j^0)/\tau=\psi(x_j)$ 只能适合精度为 $O(\tau)$ 的差分格式。若方程离散采用显式(2.19)，则第二个边界条件需要用具有二阶精度的离散方程，不妨取为

$$(u_j^1-u_j^{-1})/2\tau=\psi(x_j).$$

此时，虽然截断误差相匹配了，但必须消去 u_j^{-1}。为此，可在(2.19)式中令 $k=0$，得

$$u_j^1=r^2(u_{j+1}^0+u_{j-1}^0)+2(1-r^2)u_j^0-u_j^{-1},$$

由上面两个式子消去 u_j^{-1}，就有

$$u_j^1=\frac{r^2}{2}[\varphi(x_{j+1})+\varphi(x_{j-1})]+(1-r^2)\varphi(x_j)+\tau\psi(x_j).$$

这样，在显式计算初值问题时，解差分方程组

$$\begin{cases} u_j^{k+1}=r^2(u_{j+1}^k+u_{j-1}^k)+2(1-r^2)u_j^k-u_j^{k-1}, \\ u_j^1=\dfrac{r^2}{2}[\varphi(x_{j+1})+\varphi(x_{j-1})]+(1-r^2)\varphi(x_j)+\tau\psi(x_j), \\ u_j^0=\varphi(x_j), \quad j=\pm1,\pm2,\cdots,k=1,2,\cdots, \end{cases}$$

就可求得差分解了。

就边值问题来说，除需要同样处理初始条件外，还需处理好边界条件。将各类边界条件作如下处理：

第一类边界条件

$$\begin{cases} u(0,t)=\mu_1(t), \\ u(1,t)=\mu_2(t), \quad t>0 \end{cases}$$

可写成

$$\begin{cases} \mu_1^k=\mu_1(t_k)=u_0^k, \\ \mu_2^k=\mu_2(t_k)=u_N^k, \quad k=1,2,\cdots \end{cases}$$

第二类边界条件

$$\begin{cases} u_x(0,t)=\bar{\mu}_1(t), \\ u_x(1,t)=\bar{\mu}_2(t), \quad t>0 \end{cases}$$

可写成

$$\begin{cases} u_1^k-u_0^k=h\bar{\mu}_1(t_k), \\ u_N^k-u_{N-1}^k=h\bar{\mu}_2(t_k), \quad k=1,2,\cdots \end{cases}$$

第三类边界条件

$$\begin{cases} (a_1u+u_x)|_{x=0}=\tilde{\mu}_1(t), \\ (a_2u+u_x)|_{x=1}=\tilde{\mu}_2(t), \quad t>0 \end{cases}$$

可离散成

$$\begin{cases} u_1^k-u_0^k+a_1(t_k)hu_0^k=h\tilde{\mu}_1(t_k), \\ u_N^k-u_{N-1}^k+a_2(t_k)hu_N^k=h\tilde{\mu}_2(t_k), \quad k=1,2,\cdots. \end{cases}$$

如果继续使用显格式(2.19)来计算边值问题，就需要解由(2.19)式和已处理好的定解条件所组成的方程组。当然，此时仍存在着截断误差匹配的问题。由于上述第二、三类边界条件的离散形式都只有 $O(h)$ 的精度，同显式(2.19)不匹配，所以我们仍需仿照处理初始条件那样的办法，重新处理第二、三类边界条件。读者现已掌握了误差匹配原则，且已有能力自己处理这一问题，故不再累述。

4.3 等价方程组的差分格式

从波动方程的等价方程组(2.18)出发，构造各种差分格式也是适合的。

例如，显格式(2.19)的等价差分格式(2.20)式，就是在(2.18)式的分量形式

$$\begin{cases}\partial v/\partial t = a\partial w/\partial x; \\ \partial w/\partial t = a\partial v/\partial x\end{cases}$$

中对第一个方程取显式，对第二个方程取隐式的结果。

又如，§2.1中的L-F格式可用于波动方程，稳定性条件是$a\tau/h\leqslant 1$。§2.2中的L-W格也可用于波动方程，稳定性条件也是$a\tau/h\leqslant 1$。

波动方程的Crank-Nicholson格式

$$\frac{U_j^{k+1}-U_j^k}{\tau}+\frac{A}{2}\left(\frac{U_{j+p}^{k+1}-U_{j-p}^{k+1}}{2h}+\frac{U_{j+p}^k-U_{j-p}^k}{2h}\right)=0,$$

其中，$U=(v,w)^T$，$v=\dfrac{\partial u}{\partial t}$，$w=a\dfrac{\partial u}{\partial x}$，$p=1$ 或 $p=\dfrac{1}{2}$，是一个绝对稳定的格式，参见§1.5。

例如 $p=1/2$时，这个格式的具体形式为

$$\begin{cases}\dfrac{v_j^{k+1}-v_j^k}{\tau}=a\dfrac{w_{j+1/2}^k-w_{j-1/2}^k+w_{j+1/2}^{k+1}-w_{j-1/2}^{k+1}}{2h}, \\ \dfrac{w_{j-1/2}^{k+1}-w_{j-1/2}^k}{\tau}=a\dfrac{v_j^{k+1}-v_{j-1}^{k+1}+v_j^k-v_{j-1}^k}{2h}.\end{cases} \tag{2.21}$$

现证明这个差分格式是无条件稳定的。为此令

$v_j^k=v^k(l)e^{i\sigma x}$，$w_{j-1/2}^k=w^k(l)\exp[i\sigma(x-h/2)]$，代入(2.21)式得到传播矩阵

$$G(\sigma,\tau)=\begin{pmatrix}\dfrac{1-c^2/4}{1+c^2/4} & \dfrac{ic}{1+c^2/4} \\ \dfrac{ic}{1+c^2/4} & \dfrac{1-c^2/4}{1+c^2/4}\end{pmatrix},$$

其中，$c=2a\dfrac{\tau}{h}\sin\dfrac{\sigma h}{2}$，再求出$G$的特征值的模$|\mu|=1$，又因$G^*G$是单位阵，$G$是$U$矩阵，所以由定理1.5的3°知隐格式(2.21)是绝对稳定的。

4.4 二维波动方程的差分方法

二维波动方程

$$\partial^2 u/\partial t^2 - a^2\partial^2 u/\partial x^2 - a^2\partial^2 u/\partial y^2 = 0 \quad (a>0)$$

也可以化为二维一阶双曲型方程组

$$\partial U/\partial t + A\partial U/\partial x + B\partial U/\partial y = 0 \tag{2.22}$$

的形式，其中 $U = (v, w, z)^T$，且

$$v = \frac{\partial u}{\partial t}, \quad w = a\frac{\partial u}{\partial x}, \quad z = a\frac{\partial u}{\partial y},$$

$$A = \begin{pmatrix} 0 & -a & 0 \\ -a & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad B = \begin{pmatrix} 0 & 0 & -a \\ 0 & 0 & 0 \\ -a & 0 & 0 \end{pmatrix}.$$

和抛物型方程(组)一样，一个空间自变量的双曲型方程(组)的差分方法原则上都可以推广到两个空间自变量情形，但稳定性条件会变得更苛刻些。

例如 L-W 格式，因为

$$\partial U/\partial t = -A\partial U/\partial x - B\partial U/\partial y,$$

$$\begin{aligned}\partial^2 U/\partial t^2 &= -A\partial(\partial U/\partial t)/\partial x - B\partial(\partial U/\partial t)/\partial y \\ &= A^2\partial^2 U/\partial x^2 + (AB+BA)\partial^2 U/\partial x\partial y \\ &\quad + B^2\partial^2 U/\partial y^2,\end{aligned}$$

所以有

$$\begin{aligned}U(x,y,t+\tau) &= U(x,y,t) - \tau(A\partial U/\partial x + B\partial U/\partial y) \\ &\quad + \frac{\tau^2}{2}(A^2\partial^2 U/\partial x^2 + (AB+BA)\partial^2 U/\partial x\partial y \\ &\quad + B^2\partial^2 U/\partial y^2) + O(\tau^3).\end{aligned}$$

对上式中的偏导数全用中心差商代替，令 $r = \tau/h$，

$$T_x U_{ij} = U_{i+1,j},\ T_x^{-1}U_{ij} = U_{i-1,j},\ T_y U_{ij} = U_{i,j+1},$$

$$T_y^{-1}U_{ij} = U_{i,j-1},$$

就得 L-W 格式

$$U_{ij}^{k+1}=L_hU_{ij}^k, \tag{2.23}$$

其中差分算子

$$L_h=I-\frac{r}{2}[A(T_x-T_x^{-1})+B(T_y-T_y^{-1})]$$

$$+\frac{1}{2}r^2[A^2(T_x-2I+T_x^{-1})+B^2(T_y-2I+T_y^{-1})]$$

$$+\frac{1}{8}r^2[(AB+BA)(T_x-T_x^{-1})(T_y-T_y^{-1})].$$

用 Fourier 分析方法可以证明，当A和B都是实对称矩阵时，L-W 格式(2.23)的稳定性条件是 $r\rho(A)\leqslant 1/2$ 且 $r\rho(B)\leqslant 1/2$。这比一维情形 L-W 格式(见§1.3)的稳定性条件 $r\rho(A)\leqslant 1$ 要苛刻些。

对双曲型方程组(2.22)的分裂算法研究也是极有意义的，1968年 Strang 提出一种分裂算法，现介绍如下：

$$\begin{cases}\text{Strang 可裂格式：} U_{ij}^{k+1}=L_h(\tau)U_{ij}^k,\\ \text{其中，} r=\tau/h,\\ L_h(\tau)=L_h^x\left(\frac{\tau}{2}\right)L_h^y(\tau)L_h^x\left(\frac{\tau}{2}\right),\\ L_h^x(\tau)=I-\frac{1}{2}rA(T_x-T_x^{-1})+\frac{1}{2}r^2A^2(T_x-2I+T_x^{-1}),\\ L_h^y(\tau)=I-\frac{1}{2}rB(T_y-T_y^{-1})+\frac{1}{2}r^2B^2(T_y-2I+T_y^{-1}).\end{cases} \tag{2.24}$$

这个格式逼近 (2.22) 式且是二阶精度的。事实上，因为 $L_h^x(\tau)$和 $L_h^y(\tau)$是分别逼近

$$L^x(\tau)=I-\tau A\frac{\partial}{\partial x}+\frac{\tau^2}{2}A^2\frac{\partial^2}{\partial x^2}$$

和

$$L^y(\tau)=I-\tau B\frac{\partial}{\partial y}+\frac{\tau^2}{2}B^2\frac{\partial^2}{\partial y^2}$$

的L-W格式的差分算子(参见§1.3)，它们都是二阶精度的；又因为当$U(x,t)$足够光滑时，有

$$L^x\left(\frac{\tau}{2}\right)L^y(\tau)L^x\left(\frac{\tau}{2}\right)U(x,t)$$

$$=\left(I-\frac{\tau}{2}A\frac{\partial}{\partial x}+\frac{\tau^2}{8}A^2\frac{\partial^2}{\partial x^2}\right)\left(I-\tau B\frac{\partial}{\partial y}+\frac{\tau^2}{2}B^2\frac{\partial^2}{\partial y^2}\right)$$

$$\cdot\left(I-\frac{\tau}{2}A\frac{\partial}{\partial x}+\frac{\tau^2}{8}A^2\frac{\partial^2}{\partial x^2}\right)U(x,t)$$

$$=U-\tau\left(A\frac{\partial U}{\partial x}+B\frac{\partial U}{\partial y}\right)$$

$$+\frac{\tau^2}{2}\left[A^2\frac{\partial^2 U}{\partial x^2}+(AB+BA)\frac{\partial^2 U}{\partial x\partial y}+B^2\frac{\partial^2 U}{\partial y^2}\right]+O(\tau^3)$$

$$=U-\tau\frac{\partial U}{\partial t}+\frac{\tau^2}{2}\frac{\partial^2 U}{\partial t^2}+O(\tau^3),$$

故知Strang可裂格式是二阶精度的显格式。

Strang格式的稳定性分析也是容易的。事实上，若设G、$G_y(\tau)$、$G_x(\tau)$和$G_x\left(\frac{\tau}{2}\right)$分别是差分算子$L_h$、$L_h^y(\tau)$、$L_h^x(\tau)$和$L_h^x\left(\frac{\tau}{2}\right)$所对应的传播因子(或传播矩阵)，则有

$$|G|\leqslant\left|G_x\left(\frac{\tau}{2}\right)\right||G_y(\tau)|\left\|G_x\left(\frac{\tau}{2}\right)\right\|;$$

又因为L-W格式算子$L_h^x(\tau)$当$r\rho(A)\leqslant1$时是稳定的，且有$|G_x(\tau)|\leqslant1$，所以当$r\rho(A)\leqslant2$时有$\left\|G_x\left(\frac{\tau}{2}\right)\right\|\leqslant1$；当$r\rho(B)\leqslant1$时有$|G_y(\tau)|\leqslant1$；因此，当$r\rho(A)\leqslant2$且$r\rho(B)\leqslant1$时，Strang格式稳定。

Strang格式的稳定性条件要比二维情形的L-W格式(2.23)的稳定性条件放宽了许多，另外，Strang格式(2.24)可分裂为

三个一维显式求解，实际计算量和存贮量都很小，所以 Strang 可裂格式是很适用的。

§5 拟线性双曲型方程(组)的差分方法

拟线性双曲型方程组在空气动力学、河渠不稳定流、非线性弹性力学及松散介质力学等不少方面都会遇到，它有重要的应用价值。因此，几十年来这类方程的理论研究和数值方法研究都一直倍受人们重视，至今，新的方法和技巧仍在不断出现。

对拟线性方程(组)来说，即使其系数、非齐次项和初始条件都充分光滑，也可能出现在某些局部区域内解不唯一或解沿特征线间断的现象，这是拟线性方程(组)的特点。它反映了自然现象连续到间断的转化，空气动力学中存在的激波就是这种情形。有关间断解的差分方法将在§6中介绍，本节只讨论连续解的差分方法。

5.1 特征线方法

与线性方程一样，由拟线性方程所决定的波也是沿特征线传播的，在某点初值存在某种扰动时，这种扰动波就沿着特征线传播，这就是说，特征线就是受扰动区域和未受扰动区域的解的分界线。特别地，如果扰动使方程解中的最高阶导数有第一类间断，其它低阶导数都连续，这种解称为**弱间断解**，那么**弱间断线**就是特征线。利用这一特征理论，我们可以确定特征线的表达式。

例如，考虑一般形式的拟线性双曲型方程组

$$B(x,t,U)\partial U/\partial t + A(x,t,U)\partial U/\partial x = C(x,t,U), \tag{2.25}$$

为简明起见，其中的A和B都是2×2阶矩阵，U和C都是2×1阶矩阵，它们的元素都与U有关。

解U的全微分表达式是

$$U_x dx + U_t dt = dU, \tag{2.26}$$

因(2.25)和(2.26)式在特征线上都成立，而且由这两式联立的方

程组并不能在特征线(弱间断线)上唯一确定U_x和U_t，所以系数阵行列式为零(Cramer 法则)，这样就能确定出特征线的方向场和解在特征线上的微分关系。

具体地，(2.25)和(2.26)式的联立方程式为

$$\begin{cases} b_{11}u_{1t}+b_{12}u_{2t}+a_{11}u_{1x}+a_{12}u_{2x}=c_1, \\ b_{21}u_{1t}+b_{22}u_{2t}+a_{21}u_{1x}+a_{22}u_{2x}=c_2, \\ dtu_{1t} \qquad\qquad +dxu_{1x} \qquad\qquad =du_1, \\ \qquad dtu_{2t} \qquad\qquad +dxu_{2x}=du_2. \end{cases} \tag{2.27}$$

由(2.27)式的系数阵行列式为零，即$\begin{vmatrix} B & A \\ Idt & Idx \end{vmatrix}=0$，推出$|B-\lambda A|=0\left(\lambda=\dfrac{dt}{dx}\right)$，所以有

$$\begin{vmatrix} b_{11}-\lambda a_{11} & b_{12}-\lambda a_{12} \\ b_{21}-\lambda a_{21} & b_{22}-\lambda a_{22} \end{vmatrix}=0 \quad \left(\lambda=\frac{dt}{dx}\right). \tag{2.28}$$

这是关于λ的二次方程，它应有两个互异实根（记为λ_1和λ_2），这是由方程组的双曲型本质所决定的。

由(2.27)式的某个增广矩阵的行列式为零，例如

$$\begin{vmatrix} a_{11} & a_{12} & c_1 & b_{12} \\ a_{21} & a_{22} & c_2 & b_{22} \\ dx & 0 & du_1 & 0 \\ 0 & dx & du_2 & dt \end{vmatrix}=0,$$

对此行列式按最后一行展开，就有

$$\begin{cases} (\lambda_i|A|+E)du_1+|F|du_2+|M|dx+|N|dt=0, \qquad i=1,2 \\ |A|=\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix}, \quad |E|=\begin{vmatrix} b_{12} & a_{11} \\ b_{22} & a_{21} \end{vmatrix}, \\ |F|=\begin{vmatrix} b_{12} & a_{12} \\ b_{22} & a_{22} \end{vmatrix}, \quad |M|=\begin{vmatrix} c_1 & b_{12} \\ c_2 & b_{22} \end{vmatrix}, \quad |N|=\begin{vmatrix} a_{12} & c_1 \\ a_{22} & c_2 \end{vmatrix}. \end{cases} \tag{2.29}$$

(2.28)式反映了特征线方向场，(2.29)式则反映了解沿特征线的微分关系。

一　初值问题的特征线方法

为保证初值问题的适定性，设初值线的切线与特征线方向不同。

在求解区域内任取两点(x_j^k, t_j^k)和(x_{j+1}^k, t_{j+1}^k)，设这两点处的近似解已知，为$(u_{(1)j}^k, u_{(2)j}^k)$和$(u_{(1)j+1}^k, u_{(2)j+1}^k)$，如图2.8 *a*)。过$(x_j^k, t_j^k)$点斜率大的特征线必和过$(x_{j+1}^k, t_{j+1}^k)$点斜率小的特征线交于一点，但这个交点一般不能准确知道，可由(2.28)式的离散方程组

$$\begin{cases} t_j^{k+1} - t_j^k = \lambda_{1j}^k(x_j^k, t_j^k, u_{(1)j}^k, u_{(2)j}^k)(x_j^{k+1} - x_j^k), \\ t_j^{k+1} - t_{j+1}^k = \lambda_{2j}^k(x_{j+1}^k, t_{j+1}^k, u_{(1)j+1}^k, u_{(2)j+1}^k)(x_j^{k+1} - x_{j+1}^k), \end{cases} \tag{2.30}$$

解得(x_j^{k+1}, t_j^{k+1})并作为特征线交点的近似值。再将(2.29)式也用Euler折线法离散：

$$\begin{cases} (\lambda_{1j}^k A_j + E_j)(u_{(1)j}^{k+1} - u_{(1)j}^k) + F_j(u_{(2)j}^{k+1} - u_{(2)j}^k) \\ \qquad + M_j(x_j^{k+1} - x_j^k) + N_j(t_j^{k+1} - t_j^k) = 0, \\ (\lambda_{2j}^k A_{j+1} + E_{j+1})(u_{(1)j}^{k+1} - u_{(1)j+1}^k) + F_{j+1}(u_{(2)j}^{k+1} - u_{(2)j+1}^k) \\ \qquad + M_{j+1}(x_j^{k+1} - x_{j+1}^k) + N_{j+1}(t_j^{k+1} - t_{j+1}^k) = 0, \end{cases} \tag{2.31}$$

解出点(x_j^{k+1}, t_j^{k+1})处的近似解$u_{(1)j}^{k+1}$和$u_{(2)j}^{k+1}$，上式中$u_{(1)j}^k, u_{(2)j}^k$，λ_{1j}^k是对应于点(x_j^k, t_j^k)处的近似解和特征值，$A_{j+1}, E_{j+1}, F_{j+1}$，$M_{j+1}$和$N_{j+1}$是对应于点$(x_{j+1}^k, t_{j+1}^k)$处的相应行列式值，这些行列式表示在(2.29)式中。

因此，如图2.8*a*)所示，在知道第k层上节点近似解的情况下，可用公式(2.30)和(2.31)逐个地计算出第$k+1$层上的节点近似解，直至初值线上部决定域中解的近似值全部求出。同样也可求出初值线下部的决定域中的节点近似解。

显然，上述建立在Euler折线法基础上的特征线法是比较粗糙的，这是因为(x_j^{k+1}, t_j^{k+1})不是点(x_j^k, t_j^k)与(x_{j+1}^k, t_{j+1}^k)处的

特征线的交点，而且 $u_{(1)j}^{k+1}$ 和 $u_{(2)j}^{k+1}$ 是通过一阶精度的差分格式得到的。因此，有必要修正计算过程，以便提高数值解精度。修正办法可采用如下步骤：

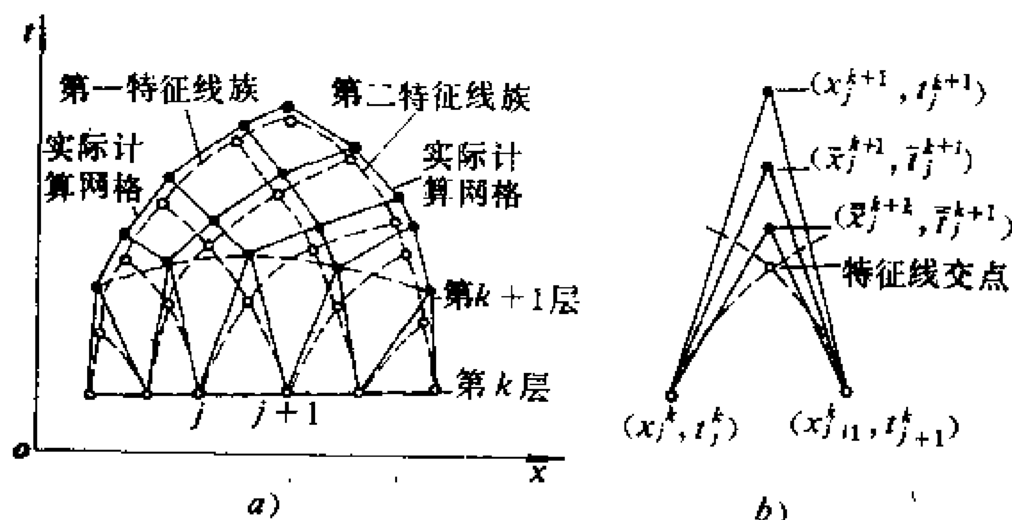


图2.8

第一步，用公式(2.30)和(2.31)计算 x_j^{k+1}、t_j^{k+1}、$u_{(1)j}^{k+1}$、$u_{(2)j}^{k+1}$、λ_{1j}^k、λ_{2j}^k、λ_{1j}^{k+1} 和 λ_{2j}^{k+1} 的数值。

第二步，修正特征方向和特征线交点。取 $\bar{\lambda}_{ij}^k=(\lambda_{ij}^k+\lambda_{ij}^{k+1})/2\ (i=1,2)$ 并用 $\bar{\lambda}_{ij}^k$ 代替(2.30)式的 $\lambda_{ij}^k\ (i=1,2)$，从而算出特征线交点的修正值 $(\bar{x}_j^{k+1},\bar{t}_j^{k+1})$。

第三步，用 $\bar{\lambda}_{ij}^k\ (i=1,2)$、$\bar{x}_j^{k+1}$ 和 $\bar{t}_j^{k+1}$ 代入(2.31)式，算出近似解修正值 $\bar{u}_{(i)j}^{k+1}\ (i=1,2)$。

显然经修正后的 $(\bar{x}_j^{k+1},\bar{t}_j^{k+1})$ 比 (x_j^{k+1},t_j^{k+1}) 更接近真实值，参见图2.8*b*)。如果有必要，还可用上述办法继续修正。一般地，改进两次就可以了。

二　Goursat 问题的特征线方法

在点 A 发出的两条特征线 $\overset{\frown}{AC}$ 和 $\overset{\frown}{AB}$ 上给定解的值，求满足此边值条件的方程组(2.25)的解。这就是 **Coursat 问题**。如图2.9所示，Goursat 问题的决定域是 $ABCD$ 曲四边形。

用特征线法解 Goursat 问题的过程如下：

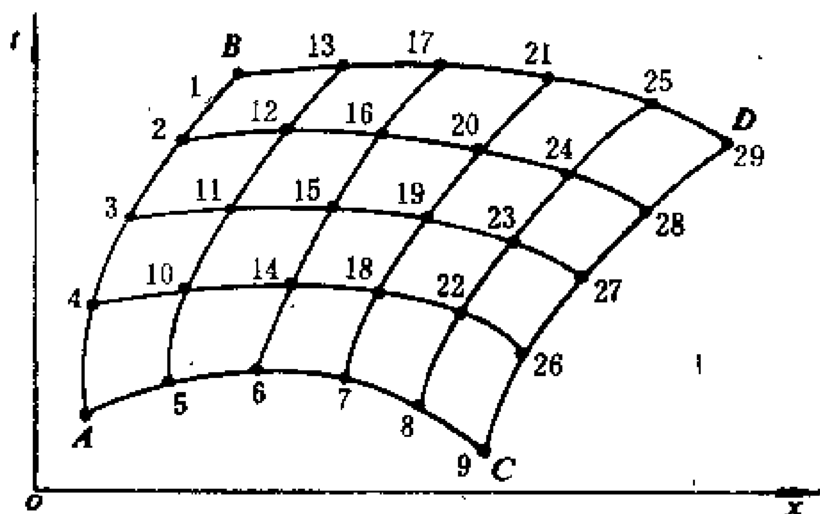


图2.9

先利用点"4"和"5"求出"10"处的近似解，再由点"3"和"10"求出"11"处的近似解，最后由点B和"12"求出"13"处的近似解。然后，将点"5"、"10"、"11"、"12"和"13"处的近似解作为初值，又可计算出另一排近似解。如此下去，直至最后求出$ABCD$曲四边形上的全部节点近似解。

三 混合边值问题的特征线方法

在$G=\{(x,t)\,|\,0<x<l,t>0\}$上求(2.25)式的解$U$，满足初始条件和适当的边值条件，使解存在唯一，这就是**混合问题**。如何给定边值条件才是恰当的，这里只作一些启发性的分析，严格的论证参见[2]、[3]。

为简明起见，设$U=(u,v)$，$\lambda_1=\lambda_1(u)$，$\lambda_2=\lambda_2(v)$，且λ和λ_2在G中不变号，那么抓住解沿特征线传播的本质，即知当特征线从左边界进入G到达右边界E点时，E点处的值与左边值有关，与右边值无关。据此道理推知，应按下述三种情况给定边值条件：

(1) 当$\lambda_1>0>\lambda_2$时，边值条件应为$u(0,t)=u_1(t)$，$v(l,t)=v_2(t)$；

(2) 当$\lambda_1>\lambda_2>0$时，边值条件应为$u(0,t)=u_1(t)$，$v(0,t)$

$=v_1(t)$；

(3) 当 $\lambda_1<\lambda_2<0$ 时，边值条件应为 $u(l,t)=u_2(t)$，$v(l,t)=v_2(t)$。

如图2.10所示，解此混合问题的特征线法应分多步进行。先由初值问题的特征线法求出 G_1 上全部节点处的数值解。当然包括 $\widehat{AC}$ 和 $\widehat{BC}$ 上的节点数值解。

再求 G_2 和 G_3 上的数值解，现以 G_2 区域为例，利用点"2"和"5"求出"6"处的数值解，利用"3"和"6"求出"7"处的数值解，…。

在获得 $\widehat{CD}$ 和 $\widehat{CE}$ 上的节点近似解后，即可用 Goursat 问题的特征线法求出 G_4 上的全部数值解。最后可求出整个决定域 $[AB]\times[0,T]$ 上的数值解。

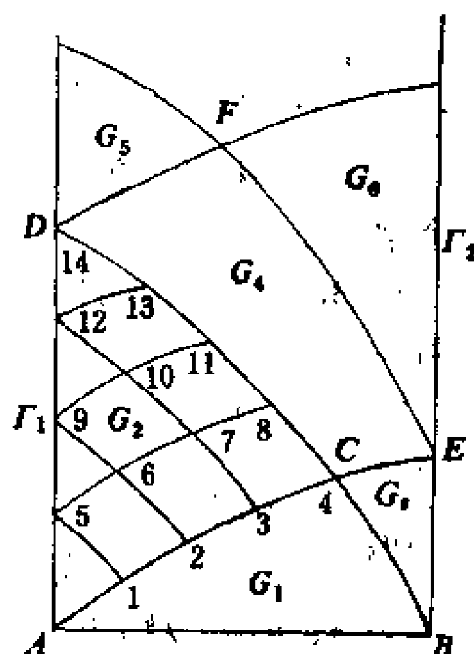


图2.10

5.2 特征差分格式

用**特征差分格式**求解拟线性方程组也是很成功的，它不但能用于解的光滑情形，而且当解出现间断时仍能较顺利地进行计算。

因特征差分格式需要考虑特征线走向和 CFL 条件，所以考虑拟线性方程组

$$\partial U/\partial t+A(U)\partial U/\partial x=C(U) \tag{2.32}$$

的特征差分格式时，必须先确定方程(2.32)的特征关系，然后再构造特征差分格式。

因为方程(2.32)是双曲型的，故 $n\times n$ 阶矩阵 $A(U)$ 有 n 个互异的实特征值，从而存在 n 个线性无关的左特征矢量 $L^i=(l_1^i,\cdots,$

l_n^i), $i=1,\cdots,n$, 使 $L^iA=\lambda_iL^i$, $\lambda_i=dx/dt$ 成立。这样，只要找出 L^i，并左乘(2.32)式两边，可得

$$L^i\partial U/\partial t+L^iA\partial U/\partial x=L^iC, \qquad i=1,\cdots,n,$$

或

$$\sum_{q=1}^{n} l_q^i\left(\frac{\partial u_q}{\partial t}+\lambda_i\frac{\partial u_q}{\partial x}\right)=\sum_{q=1}^{n} l_q^i c_q, \qquad i=1,\cdots,n, \tag{2.33}$$

或

$$\sum_{q=1}^{n} l_q^i\left(\frac{du_q}{dt}\right)_{\tau_i}=\sum_{q=1}^{n} l_q^i c_q, \qquad i=1,\cdots,n,\ \tau_i \text{表示特征方向.}$$

(2.33)式称为**特征关系式**，它表明特征线上解的微分关系是一个常微分方程组.

(2.33)式的特征方向 λ_i 非常明显，可用特征关系式直接构造特征差分格式：

$$\begin{cases}\sum_{q=1}^{n} l_{q,j}^{i,k}\left(\frac{u_{q,j}^{k+1}-u_{q,j}^{k}}{\tau}+\lambda_{i,j}^{k}\Delta u_{q,j}^{k}\right)=\sum_{q=1}^{n} l_{q,j}^{i,k}c_{q,j}^{k},\\ \Delta u_{q,j}^{k}=\begin{cases}(u_{q,j}^{k}-u_{q,j-1}^{k})/h, & \text{当 } \lambda_{i,j}^{k}\geqslant 0 \text{时},\\ (u_{q,j+1}^{k}-u_{q,j}^{k})/h, & \text{当 } \lambda_{i,j}^{k}<0 \text{时},\end{cases}\end{cases} \tag{2.34}$$

其中 $l_{q,j}^{i,k}$、$u_{q,j}^{k}$、$\lambda_{i,j}^{k}$ 和 $c_{q,j}^{k}$ 分别表示(2.33)式中的 l_q^i、u_q、λ_i 和 c_q 在 (x_j,t_k) 点处的值.

特征差分格式(2.34)式中已考虑了特征线走向，但还必须满足 CFL 条件，为此要求

$$\frac{\tau}{h}|\lambda_i|\leqslant 1 \quad \text{或 } \tau/h\leqslant 1/\max_{q,j,k}|\lambda_{q,j}^{k}|.$$

值得注意的是，拟线性差分格式的稳定性讨论较复杂，而且 Lax 等价定理此时不成立，上式只是仿照冻结系数法的结果，它并非是严格的稳定性充分条件.

§6 守恒型双曲方程(组)的广义解及其差分方法

双曲型方程(组)的解常常出现间断，这种间断可以分为两大

类。一类间断是：解本身连续，但它的导数间断，这类间断称为**弱间断**。例如流体力学中稀疏波的波头就是一种弱间断。弱间断情形的差分方法已在§5.1中讨论过。另一类间断是解本身间断，称为**强间断**。例如冲击波等就是强间断。解的这些间断特性是双曲型方程(组)区别于其它类型方程(组)的又一个重要特性。对于椭圆型方程(组)和抛物型方程(组)来说，虽然边界条件或初始条件出现间断情形，但是在离开间断发生点的区域内部不远处，这两种类型定解问题的解很快就变得光滑了。然而双曲型方程(组)的定解问题却不同，初始条件的间断扰动会直接传播进定解区域内部。对于简单的对流方程(2.3)和线性双曲型方程组((2.7)式及(2.14)式)来说，这一表现是非常清楚的，而且特征线就是解的间断线。

对于拟线性双曲型方程(组)(2.25)式来说，解的间断表现得更为突出和复杂。不仅初始条件的间断扰动会直接传播到定解区域内部，(2.25)式中的系数阵 $B(x,t,U)$、$A(x,t,U)$和$C(x,t,U)$的元素的间断也会造成解的间断；而且即使 $B(x,t,U)$、$A(x,t,U)$和 $C(x,t,U)$中的元素和初始条件都充分光滑，也只能保证(2.25)式有局部光滑解，在大范围内解会沿着某些曲线(面)间断，甚至可能出现在某些局部区域内解不唯一存在的情形。

例如拟线性双曲型方程

$$\begin{cases} \dfrac{\partial u}{\partial t}+u\dfrac{\partial u}{\partial t}=0 \quad (t>0,\ -\infty<x<+\infty), \\ u(x,0)=v(x) \quad (-\infty<x<+\infty), \end{cases}$$

其特征线方向为 $dx/dt=u(x,t)$，在特征线上解满足微分关系 $du/dt=0$，即解沿特征线是常数。过$(x_0,0)$点的特征线方程是

$$x=x_0+u(x_0,0)t=x_0+v(x_0)t.$$

若初始条件分别取为

$$v(x)=\begin{cases}1, & x>0;\\ -1, & x\leqslant 0,\end{cases}\quad \text{或}\quad v(x)=\begin{cases}-1, & x>0;\\ 1, & x\leqslant 0,\end{cases}$$

则方程解沿特征线传播情况分别如图2.11a)和图2.11b)所示。在

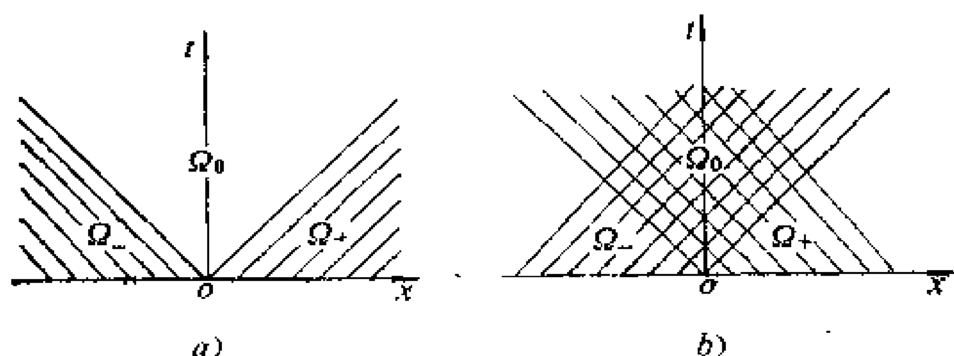


图2.11

Ω_- 和 Ω_+ 部分，解是唯一的；在 Ω_0 部分无经典解或解不唯一；Ω_- 与 Ω_0 及 Ω_+ 与 Ω_0 的交界线都是特征线，它们都是解的间断线。

拟线性双曲型方程的解发生间断、无解和多值现象不一定是由于初始条件存在间断造成的，而是由于双曲型方程解沿着特征线传播这一特点所至。既使初值函数充分光滑，也会出现特征线相交或不相交的区域。例如初值问题

$$\frac{\partial u}{\partial t}+u\frac{\partial u}{\partial x}=0,\ u(x,0)=-\mathrm{th}\frac{x}{2\varepsilon},\ \ \varepsilon>0$$

的初值函数$\left(\mathrm{th}\,y=\frac{e^y-e^{-y}}{e^y+e^{-y}}\right)$充分光滑，但是该问题的光滑解仅当 $0\leqslant t<2\varepsilon$ 时局部存在。事实上，由于特征线方向$\left.\frac{dx}{dt}\right|_{t=0}=$ $u(x,t)|_{t=0}=u(x,0)$，故经过 $(x_0,0)$ 和 $(-x_0,0)$ 两点的特征线分别为

$$x=-\mathrm{th}\frac{x_0}{2\varepsilon}t+x_0 \text{ 和 } x=\mathrm{th}\frac{x_0}{2\varepsilon}-x_0.$$

这两条特征线相交于 $P\left(0,\ x_0/\mathrm{th}\frac{x_0}{2\varepsilon}\right)$，这表明 u 在 P 点处有两

个值 $u=\pm\operatorname{th}\frac{x_0}{2\varepsilon}$，因而出现多值性。当 $x_0\to 0$ 时，用罗必达法则可知点P趋于点$(0,2\varepsilon)$。所以单值的光滑解仅当 $0\leqslant t<2\varepsilon$ 时局部存在，而当 $t\geqslant 2\varepsilon$ 时，由于多值性而出现间断。

拟线性双曲型方程的解存在着复杂的间断情形，这不仅是理论研究的对象，而且也是实际问题中极需重视的普遍问题。例如，Courant 等人早在1948年就详细分析了一维等熵流，他们在一个无限长的绝热管内充满气体，管的左端用活塞封住，活塞在时刻 $t=0$ 从点 $x=0$ 由静止开始并以小的正加速度连续地向管内移动，活塞的运动规律是 $x=X(t)$。此问题的数学模型是一维等熵流动问题

$$\begin{cases}\dfrac{\partial\rho}{\partial t}+u\dfrac{\partial\rho}{\partial x}+\rho\dfrac{\partial u}{\partial x}=0;\\[2mm]\dfrac{\partial u}{\partial t}+u\dfrac{\partial u}{\partial x}+\dfrac{a^2}{\rho}\dfrac{\partial\rho}{\partial x}=0;\\[2mm]u(x,0)=u_0(x);\\ \rho(x,0)=\rho_0(x),\end{cases}$$

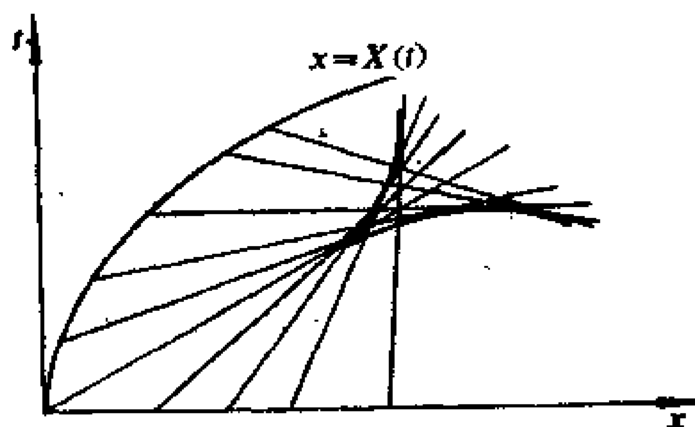


图2.12

其中，u 表示气体流动速度，ρ 表示气体密度，$a=a(\rho)$ 是声速。因为在任何时刻 t，活塞处的气体运动速度等于活塞的速度，即

$u=\frac{dx}{dt}$，求解区域就是图2.12中所示的 $x=X(t)$ 曲线和 x 轴所围成的区域．此时有两个特征 $dx/dt=u\pm a$。在给定的条件下，沿求解区域的边界线有两族特征线(直线)，它们形成一个有尖点的包络，而且在包络围成的区域内，通过每点有三条（而不是一条)特征线．这说明解沿包络间断，且在包络线围成的区域内部解是不唯一的．这一实验表明，经过一段时间后，流动中就产生了一个不连续的扰动，这种激震现象称作**激波**．激波在流动（特别是高速流动）的理论中是非常重要的．Courant等人的这一实验对于流体力学中的许多问题都是有典型意义的．

因此，为了研究双曲型方程(组)的间断解，需要从理论和实践两个方面研究下面几个问题．

(1) 由于在解的间断处，偏微分方程已失去了通常的意义，方程的各阶偏导数可能不再连续，所以有必要把**经典解**的意义加以推广而研究**广义解**(**弱解**)．

(2) 在实际问题中广义解应该是唯一存在的，这种解称为**物理解**，因此物理解应得到合理的数学描述．

(3) 研究物理解所具有的数学性质，这方面的理论研究将对数值求解双曲型问题有指导意义．

(4) 间断解的间断位置很难确定，因此有必要研究和构造种种通用的差分格式，称为**一致差分格式**，使用这种差分格式不用顾及特征线的走向，也不用顾及解的间断位置，差分解能同时有效地描述光滑解和间断解．

(5) 研究一致差分格式对间断解的逼近效果．

(6) 研究一致差分格式的收敛性．

本节将从这6个方面讨论具有散度形式的**守恒型双曲方程**

$$\frac{\partial u}{\partial t}+\frac{\partial f(u)}{\partial x}=0,\qquad (x,t)\in R\times(0,T]=D \tag{2.35a}$$

和**守恒型双曲方程型**

$$\frac{\partial U}{\partial t}+\frac{\partial F(U)}{\partial x}=0, \quad (x,t)\in R\times(0,T] \qquad (2.35b)$$

的差分方法，其中 $U(x,t)=(u_1(x,t),u_2(x,t),\cdots,u_n(x,t))^T$。因为这类方程(组)反映了能量守恒律，所以它是有普遍意义的。若令(2.35b)式中 $\partial F(U)/\partial U=A(U)$，其中 $A(U)$ 是Jacobi矩阵，则方程组就是形如(2.32)式的拟线性双曲型方程组；另外，(2.35a)式是(2.35b)式的特殊情形。

6.1 广义解、物理解和熵条件

先对单个守恒型双曲方程 (2.35a) 讨论广义解的定义。大家知道，Green 公式

$$\int_{D_1}\left(\frac{\partial u}{\partial t}+\frac{\partial f(u)}{\partial x}\right)dxdt=\int_{\partial D_1}(f(u)dt-udx), \quad D_1\subset D$$

把区域积分化为边界积分，使得方程(2.35a)解的光滑性要求变弱。所以，只要注意Green公式所需要的条件，就可利用Green公式推广古典解的概念。

定义 1 分片光滑函数 $u(x,t)$ 称为方程 (2.35a) 在定解区域D上的**广义解(弱解)**，如果对任一连同其内部 D_1 都属于D的按段光滑的简单闭曲线 Γ，恒有

$$\int_{\Gamma}(f(u)dt-udx)=0. \qquad (2.36)$$

显然，这样定义的广义解推广了古典解的概念；广义解允许在某些曲线上有第一类间断；且当 $u\in C^1$ 时，广义解和古典解一致。

现进一步讨论定义 1 中的广义解是如何描述解在间断线两旁的变化规律的。为此，设 γ: $x=x(t)$ 是解的间断线，由分段光滑闭曲线 Γ 围成的区域 $D_1\cup D_2$ 包含着 γ，为简单起见，设 $D_1\cup D_2$ 是以 $P\in\gamma$ 为圆心的圆，Γ 是圆周，参见图2.13。

若记 u_- 和 u_+ 是 u 分别从左侧和右侧趋于间断线 γ 的极限，

且让 D_1 的左边界和D_2的右边界分别趋于γ_- 和γ_+，则由Green公式和(2.36)式有

$$\int_{\gamma_-}(f(u_-)dt-u_-dx)$$

$$=\int_{\gamma_+}(f(u_+)dt-u_+dx),$$

从而有

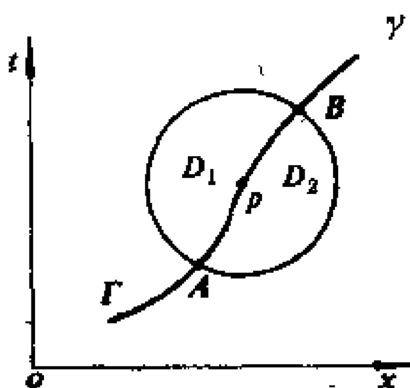


图2.13

$$\frac{dx}{dt}=\frac{f(u_+)-f(u_-)}{u_+-u_-}. \tag{2.37}$$

这就是分片可微解的**间断关系**，又称作**跳跃条件**、**连接条件**或**内边界条件**，在气体动力学中它称为**Rankine-Hugoniot条件**。由于 $f(u)$可表示流量或能量，因此间断关系(2.37)式表明：解发生间断时，流量或能量将发生激剧变化，其变化率(**间断速度**)刚好等于间断线斜率的倒数。

上述推导可以顺逆方向进行，所以有与定义 1 等价的另一个广义解定义。

定义 2 在分片光滑区 $u(x,t)$是古典解，即满足微分方程(2.35a)，而在 $u(x,t)$的间断线 γ 上满足解的间断关系(2.37)式，称这样的解 $u(x,t)$是**广义解**。

也可用下面办法定义广义解（参见第二篇中关于椭圆型方程广义解的定义），用试验函数 $\varphi(x,t)$ 乘 (2.35a)式两边且积分，就有

$$\int_D\left(\frac{\partial u}{\partial t}+\frac{\partial f}{\partial x}\right)\varphi dxdt$$

$$=\int_{-\infty}^{+\infty}[u\varphi]_{t=0}^{t=T}dx-\int_D u\frac{\partial\varphi}{\partial t}dxdt$$

$$+\int_0^T[f\varphi]_{x=-\infty}^{x=+\infty}dt-\int_D f\frac{\partial\varphi}{\partial x}dxdt=0,$$

为了能同时描述区域内部和初值线上的情形，设试验函数$\varphi(x,t)$由两种类型的函数构成。第一类：φ是具有**紧致支集**的试验函数，即$\varphi(x,t)$光滑可微，φ在区域D外（包括区域D的边界∂D上）恒等于零，区域D称为φ的**支集**，这样的函数类用E_φ表示。这样，上述积分就能描述广义解在区域D内所满足的关系

$$\int_{t>0}\left[u\frac{\partial\varphi}{\partial t}+f\frac{\partial\varphi}{\partial x}\right]dxdt=0,\qquad \forall\varphi\in E_\varphi. \tag{2.38}$$

第二类：$\varphi(x,t)$光滑可数，它在区域D以及∂D上除了和$t=0$重合的一段以外恒等于零，记这样的函数类为E，显然$E_\varphi\subset E$。这样，上述积分就能描述在区域内部连同初值线上微分方程解所满足的关系

$$\int_{t>0}\left[u\frac{\partial\varphi}{\partial t}+f\frac{\partial\varphi}{\partial x}\right]dxdt+\int_{-\infty}^{+\infty}\varphi(x,0)u(x,0)dx=0,$$
$$\forall\varphi\in E. \tag{2.39}$$

这样，我们可作出广义解的另一定义。

定义 3 若(2.38)式和(2.39)式成立，则称$u(x,t)$是方程(2.35a)满足初始条件$u(x,t)|_{t=0}=u(x,0)$的Cauchy问题的广义解。

可以证明定义1、定义2和定义3是等价的，为此只要证明定义2与定义3等价就行了。

设D中含有一条间断线γ：$x=x(t)$，如图2.13所示，设u在D_-和D_+中连续可微，则由(2.38)式可得到(3.35a)式，且有

$$\int_{t>0}\left(u\frac{\partial\varphi}{\partial t}+f\frac{\partial\varphi}{\partial x}\right)dxdt$$
$$=\int_{D_-}\left(u\frac{\partial\varphi}{\partial t}+f\frac{\partial\varphi}{\partial x}\right)dxdt+\int_{D_+}\left(u\frac{\partial\varphi}{\partial t}+f\frac{\partial\varphi}{\partial x}\right)dxdt$$
$$=-\int_{\partial D_-}(udx-fdt)\varphi-\int_{\partial D_+}(udx-fdt)\varphi$$

$$= -\int_{\gamma_-}[u_- dx - f(u_-)dt]\varphi$$

$$-\int_{\gamma_+}[u_+ dx - f(u_+)dt]\varphi$$

$$= \int_{\gamma}[(u_+ - u_-)dx - (f(u_+) - f(u_-))dt]\varphi = 0,$$

$$\forall \varphi \in E_\varphi.$$

这样由定义 3 就推得，在间断线上解满足间断条件(2.37)。如果 D 含有 $u(x,t)$ 的几条间断线，也可用类似的办法推得这一结果。

同样，由定义 2 也易推得定义 3．

上面给出了三种广义解的定义，它们是相互等价的．但需要指出的是，这样定义的广义解是不唯一的．例如对 Cauchy 问题

$$\begin{cases} \dfrac{\partial u}{\partial t} + \dfrac{\partial (u^2/2)}{\partial x} = 0, & x \in (-\infty, +\infty), t > 0, \\ u(x,0) = \begin{cases} -1, & x \leqslant 0, \\ 1, & x > 0, \end{cases} \end{cases}$$

可以构造

$$u(x,t) = \begin{cases} -1, & x \leqslant -t, \\ \dfrac{x}{t}, & -1 < \dfrac{x}{t} \leqslant 1, \\ 1, & x > t. \end{cases}$$

作为连续的广义解，也可以构造($|a| \leqslant 1$)

$$u_a(x,t) = \begin{cases} -1, & x < \dfrac{a-1}{2}t, \\ a, & \dfrac{a-1}{2}t \leqslant x < 0, \\ -a, & 0 \leqslant x \leqslant \dfrac{1-a}{2}t, \\ 1, & x > \dfrac{1-a}{2}t, \end{cases}$$

作为分片光滑的广义解。不唯一的广义解是没有物理意义的。看来，必须研究附加某些物理条件限制的唯一的广义解，这称为**物理解**。物理解是符合实际物理过程的，因此研究物理解是研究守恒型双曲方程(组)差分方法的核心问题。

现已证明，满足熵条件的广义解是唯一的，即是物理解，为简单起见，现就一维守恒型双曲方程的物理解叙述之。

设 $u(x,t)$ 是定义在 $t\geqslant 0$ 上半平面的，它是除了在有限条光滑曲线上间断外的连续可微函数。将这样的函数集合记作 K。

定理2.1 守恒型双曲方程 (2.35a) 满足初始条件 $u(x,0)=v(x)\ (-\infty<x<+\infty)$ 的广义解 $u(x,t)$，若在其间断线上满足

$$\frac{f(u_-)-f(w)}{u_--w}\geqslant\frac{f(u_+)-f(u_-)}{u_+-u_-}$$

$$\geqslant\frac{f(u_+)-f(w)}{u_+-w},\quad \forall w\in I, \tag{2.40}$$

其中 $I=(\min\{u_-,u_+\},\ \max\{u_-,u_+\})$，则这样的广义解在函数类 K 中是唯一的。

不等式(2.40)称为**熵条件**，用(OE)表示。

证明 为了简单起见，我们只就初值 $v(x)$ 是具有紧致支集的有界函数和广义解也是有界的情况进行讨论。设 $v(x)$ 的支集包含在区间 $[-x_0,x_0]$ 之中，弱解的界为 M_0，即

$$|u(x,t)|<M_0.$$

令 $\sup\limits_{|u|<M_0}\left|\dfrac{df(u)}{du}\right|=A$，于是对任意的 $t>0$，$u(x,t)$ 作为 x 的函数也是具有紧致支集的函数，其支集包含在区间 $[-(x_0+At),x_0+At]$ 之中。

设 $u(x,t)$ 和 $w(x,t)$ 都属于 K，它们都是方程 (2.35a) Cauchy 问题的两个弱解，则这两个函数之差在 L_1 中的模可以看成是 t 的函数，记为

$$H(t)=\|u-w\|_{L_1}=\int_{-\infty}^{+\infty}|u(x,t)-w(x,t)|dx,$$

显然有 $H(0)=0$，$H(t)\geqslant 0$。函数 $H(t)$ 可以写成

$$H(t)=\sum_{m=0}^{M-1}(-1)^m\int_{y_m(t)}^{y_{m+1}(t)}[u(x,t)-w(x,t)]dx$$

的形式，其中 $x=y_0(t)$ 表示直线 $x=-x_0-At$，$x=y_M(t)$ 表示直线 $x=x_0+At$，而 $x=y_m(t)\ (m=1,2,\cdots,M-1)$ 是这样一些曲线，函数 $u(x,t)-w(x,t)$ 在 $y_m(t)$ 的两侧异号，即在 $y_m(t)<x<y_{m+1}(t)$ 中

$$|u(x,t)-w(x,t)|=(-1)^m[u(x,t)-w(x,t)].$$

分别讨论以下两种情况：

(1) 在 $y_m(t)$ 上 $u(x,t)$ 和 $w(x,t)$ 都是连续的。这时有

$$u(y_m(t),t)=w(y_m(t),t);$$

(2) 在 $y_m(t)$ 上 $u(x,t)$ 间断，而 $w(x,t)$ 连续。仍然用 u_- 和 u_+ 分别表示 u 在间断线左右两侧的极限值，这时又存在两种情况：

ⓐ$u_->w>u_+$，　m 为奇数；

ⓑ$u_+>w>u_-$，　m 为偶数。

由于 $x=y_m(t)$ 是 $u(x,t)$ 的间断线，所以一定有

$$\frac{dy_m}{dt}=\frac{f(u_+)-f(u_-)}{u_+-u_-}$$

(至于其它情况，例如在 y_m 上 $u(x,t)$ 和 $w(x,t)$ 都间断等等，都可仿照讨论(2)的方法加以证明)。

现在要求出 $H(t)$ 对 t 的导数：

$$\frac{dH}{dt}=\sum_{m=0}^{M-1}(-1)^m\frac{d}{dt}\int_{y_m(t)}^{y_{m+1}(t)}(u-w)dx.$$

由于 u 和 w 都属于K，所以在 $y_m(t)<x<y_{m+1}(t)$ 之间最多只可能有有限多个间断线。首先考虑在 $y_m(t)<x<y_{m+1}(t)$ 之间，u 在 $x=z(t)$ 上间断，在其两侧 u 的极限值取 u_-^* 和 u_+^*，而 w 假定是连续的。那么

$$\frac{d}{dt}\int_{y_m(t)}^{y_{m+1}(t)}(u-w)dx$$

$$= \frac{d}{dt}\left[\int_{y_m(t)}^{z(t)} (u-w)\,dx + \int_{z(t)}^{y_{m+1}(t)} (u-w)\,dx\right]$$

$$= \int_{y_m(t)}^{z(t)} \frac{\partial}{\partial t}(u-w)\,dx + \int_{z(t)}^{y_{m+1}(t)} \frac{\partial}{\partial t}(u-w)\,dx$$

$$+ (u-w)\frac{dx}{dt}\Big|_{x=y_m(t)}^{x=y_{m+1}(t)} - (u_+^z - u_-^z)\frac{dz}{dt}$$

$$= \int_{y_m(t)}^{z(t)} -\frac{\partial}{\partial x}[f(u)-f(w)]dx$$

$$+ \int_{z(t)}^{y_{m+1}(t)} -\frac{\partial}{\partial x}[f(u)-f(w)]dx$$

$$+ (u-w)\frac{dx}{dt}\Big|_{x=y_m(t)}^{x=y_{m+1}(t)} - (u_+^z - u_-^z)\frac{dz}{dt}$$

$$= -\left[f(u)-f(w)-(u-w)\frac{dx}{dt}\right]_{x=y_m(t)}^{x=y_{m+1}(t)}$$

$$+ f(u_+^z) - f(u_-^z) - (u_+^z - u_-^z)\frac{dz}{dt}.$$

故不论 u 和 w 在 $y_m(t)<x<y_{m+1}(t)$ 之间有多少根间断线，总有

$$\frac{dH}{dt} = \sum_{m=0}^{M-1}(-1)^{m+1}\left[f(u)-f(w)-(u-w)\frac{dx}{dt}\right]_{x=y_m(t)}^{x=y_{m+1}(t)}$$

$$= \sum_{m=0}^{M-1}(-1)^m\left\{\left[f(u)-f(w)-(u-w)\frac{dx}{dt}\right]_{x=y_m(t)-0}\right.$$

$$\left.+\left[f(u)-f(w)-(u-w)\frac{dx}{dt}\right]_{x=y_m(t)+0}\right\}.$$

令

$$J_m = (-1)^m\left\{\left[f(u)-f(w)-(u-w)\frac{dx}{dt}\right]_{x=y_m(t)-0}\right.$$

$$+\left[f(u)-f(w)-(u-w)\frac{dx}{dt}\right]_{x=y_m(t)+0}\Big\}.$$

对于情形(1)，由于 $u(y_m(t),t)=w(y_m(t),t)$，故得到 $J_m=0$。在讨论情形(2)之前，先将 J_m 的形式变换成

$$J_m=(-1)^m 2(u_+-w)\left[\frac{f(u_+)-f(w)}{u_+-w}-\frac{f(u_+)-f(u_-)}{u_+-u_-}\right];$$

或者，还可变换为

$$J_m=(-1)^m 2(w-u_-)\left[\frac{f(u_+)-f(u_-)}{u_+-u_-}-\frac{f(w)-f(u_-)}{w-u_-}\right].$$

因此对于情形(2)ⓐ，即当 m 为奇数时，有

$$J_m=2(w-u_+)\left[\frac{f(u_+)-f(w)}{u_+-w}-\frac{f(u_+)-f(u_-)}{u_+-u_-}\right],$$

或

$$J_m=2(u_--w)\left[\frac{f(u_+)-f(u_-)}{u_+-u_-}-\frac{f(w)-f(u_-)}{w-u_-}\right].$$

对于情形(2)ⓑ，即当 m 为偶数时，有

$$J_m=2(u_+-w)\left[\frac{f(u_+)-f(w)}{u_+-w}-\frac{f(u_+)-f(u_-)}{u_+-u_-}\right],$$

或

$$J_m=2(w-u_-)\left[\frac{f(u_+)-f(u_-)}{u_+-u_-}-\frac{f(w)-f(u_-)}{w-u_-}\right].$$

如果

$$\frac{f(u_+)-f(w)}{u_+-w}\leqslant\frac{f(u_+)-f(u_-)}{u_+-u_-},\quad \forall w\in I, \tag{2.41a}$$

或它的等价形式

$$\frac{f(u_+)-f(u_-)}{u_+-u_-}\leqslant\frac{f(w)-f(u_-)}{w-u_+},\quad \forall w\in I \tag{2.41b}$$

成立，则 $J_m \leqslant 0$。如果对于每一个 m，当 u 间断时，都有上面两个不等式成立，则得 $dH/dt \leqslant 0$。从H的定义和 $H(0)=0$ 可推出 $H(t)=0(t\geqslant 0)$。这样，条件(2.41a)和(2.41b)就保证了在函数类K中，广义解在 L_1 模的意义下是唯一的。而条件(2.41a)和(2.41b)正是熵条件(2.40)。 □

定理2.1说明：对单个守恒型双曲方程来说，满足熵条件的广义解是物理解。物理解不仅是唯一的，而且在 L_1 模意义下是连续依赖于初始条件的，即是稳定的。这是因为如果 $H(0)<\varepsilon$，则由于熵条件(OE)成立，就有 $\frac{dH}{dt}\leqslant 0$，故得到

$$H(t)<\varepsilon,\quad t>0.$$

熵条件(OE)无论对理论研究，还是对数值研究守恒型双曲方程来说都是非常重要的。特别地，若广义解在间断处适合熵条件(OE)中的严格不等关系，即

$$\frac{f(u_-)-f(w)}{u_- - w} > \frac{f(u_+)-f(u_-)}{u_+ - u_-} > \frac{f(u_+)-f(w)}{u_+ - w},$$

$$\forall w\in I,$$

则把这类间断称为**激波**。$u_+ - u_-$ 称为**激波强度**。激波在流体的高速流动中是经常发生的。若熵条件(OE)中的等号对 $\forall w\in I$成立，此类间断称为**接触间断**。为了加深理解，下面对熵条件(OE)作一些解释和分析。

在$(u,f(u))$平面上考察曲线 $f(u)$。当 $u_+<w<u_-$ 时，(OE)可改写成

$$f(w)\leqslant\frac{u_- - w}{u_- - u_+}f(u_+)+\frac{w-u_+}{u_- - u_+}f(u_-). \tag{2.42a}$$

这正表示曲线 $f(u)$ 在区间 $I=(u_+,u_-)$ 上位于连接$(u_+,f(u_+))$和$(u_-,f(u_-))$两点的直线的下面（参见图2.14）。而当 $u_-<w<u_+$ 时，(OE)可以写成

$$f(w) \geqslant \frac{u_+ - w}{u_+ - u_-} f(u_-) + \frac{w - u_-}{u_+ - u_-} f(u_+). \qquad (2.42b)$$

这表示曲线 $f(u)$ 在区间 $I=(u_-,u_+)$ 上位于连接 $(u_-,f(u_-))$ 和 $(u_+,f(u_+))$ 两点的直线的上方(参见图2.15)。

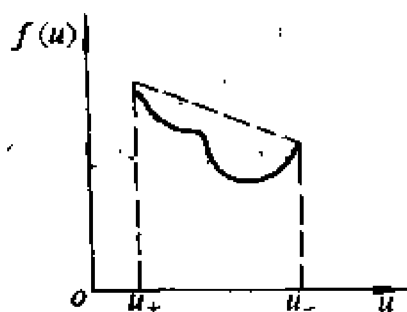


图2.14

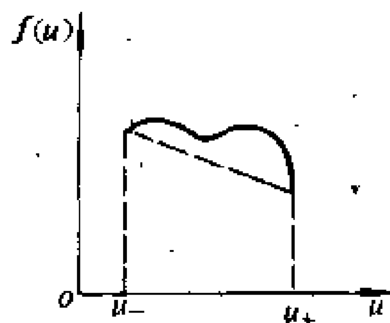


图2.15

熵条件中的 $\dfrac{f(u_+)-f(u_-)}{u_+-u_-}$ 是间断传播速度，用 S 表示。这就是 R-H 条件(2.37)式，它反映了解的间断处(激波处)某种能量 $f(u)$ 的急剧变化的变化率。若在(2.40)式左边令 $w\to u_-$，同时在(2.40)式右边令 $w\to u_+$，则得到

$$\frac{\partial f(u_-)}{\partial u} \geqslant S \geqslant \frac{\partial f(u_+)}{\partial u}. \qquad (2.43)$$

若将 $f(u)$ 视为某种能量，则上式反映了在间断线左边、间断线上和间断线右边这种能量变化率之关系；若注意到 $\dfrac{dx}{dt}=\dfrac{\partial f(u)}{\partial u}$，则上式可理解为间断线两侧的特征线都是走向间断的。不等式(2.43)是(OE)的一个直接推论，有时也当作熵条件，记作(CE)。

当 f 二阶连续可微并且 $f''(u)$ 在区间 I 上不变号时，熵条件特别简单。如果 $f''(u)<0$，这时 $f'(u)$ 是单调减的，从(CE)就推出 $u_-<u_+$；又由于 $f''(u)<0$，$f(u)$ 曲线在 $[u_-,u_+]$ 上是向上凸的，故对于任意的 $w\in I$，(2.42b)式成立，即(OE)成立；因

此当 $f''(u)<0$ 时，就可推出

$$(\mathrm{OE})\to(\mathrm{CE})\to u_-<u_+\to(\mathrm{OE}),$$

这样就得到一组等价的熵条件。同时，当 $f''(u)>0$ 时，同样可依次推出

$$(\mathrm{OE})\to(\mathrm{CE})\to u_+<u_-\to(\mathrm{OE}),$$

也得到一组等价的熵条件。

人们还可以用**粘性消失法**研究广义解和物理解。大家知道，激波的产生与流体的粘性(内摩擦)有密切关系，但(2.35a)式作为理想流体力学方程却忽略了分子的摩擦，即没有考虑粘性。一个很自然的想法是，在(2.35a)式中增加一个带小参数 $\varepsilon>0$ 的**粘性项**，使其成为一个拟线性抛物型方程：

$$\frac{\partial u_\varepsilon}{\partial t}+\frac{\partial f(u_\varepsilon)}{\partial x}=\varepsilon\frac{\partial^2 u_\varepsilon}{\partial x^2},\qquad \varepsilon>0. \tag{2.44}$$

显然，当 $\varepsilon\to0$ 时，(2.44)式的古典解 u_ε 的极限就是(2.35a)式的广义解。那么，由(2.44)式定义的广义解与前面定义的几种广义解是否一致？它是否能确定物理解？

定理 2.2 当 $\varepsilon\to0$ 时，如果方程(2.44)满足初始条件

$$u_\varepsilon(x,0)=v(x) \tag{2.45}$$

的解 $u_\varepsilon(x,t)$ 一致有界且几乎处处收敛到分片连续可微的函数 $u(x,t)$，则 $u(x,t)$ 是方程(2.35a)连同初始条件 $u(x,0)=v(x)$ 的物理解。

证明 为证明简单起见，设 $f(u)$ 是充分光滑的，且(2.44)式和(2.45)式存在唯一连续可微的解 $u_\varepsilon(x,t)$。

容易看出，当 $\varepsilon\to0$ 时，u_ε 和 $f(u_\varepsilon)$ 是弱收敛于 $u(x,t)$ 和 $f(u)$ 的。

用任定的 $\varphi(x,t)\in E$ 乘(2.44)式两端，且在 $t\geq0$ 上积分，就有

$$\int_{t>0}\left(\frac{\partial u_\varepsilon}{\partial t}\varphi+\frac{\partial f(u_\varepsilon)}{\partial x}\varphi\right)dx\,dt=\int_{t>0}\varepsilon\frac{\partial^2 u_\varepsilon}{\partial x^2}dx\,dt.$$

由于 u_ε 连续可微，所以将左边改写后上式变为

$$\int_{t>0}\left\{\frac{\partial[u_\varepsilon\varphi]}{\partial t}+\frac{\partial[f(u_\varepsilon)\varphi]}{\partial x}-\left[u_\varepsilon\frac{\partial\varphi}{\partial t}+f(u_\varepsilon)\frac{\partial\varphi}{\partial x}\right]\right\}dxdt$$

$$=\int_{t>0}\varepsilon\frac{\partial^2 u_\varepsilon}{\partial x^2}\varphi dxdt,$$

再作分部积分，便有

$$-\int_{-\infty}^{\infty}u_\varepsilon(x,0)\varphi(x,0)dx-\int_{t>0}\left[u_\varepsilon\frac{\partial\varphi}{\partial t}+f(u_\varepsilon)\frac{\partial\varphi}{\partial x}\right]dxdt$$

$$=\varepsilon\int_{t>0}u_\varepsilon\frac{\partial^2\varphi}{\partial x^2}dxdt.$$

当 $\varepsilon\to 0$ 时，就得到

$$\int_{t>0}\left[u\frac{\partial\varphi}{\partial t}+f(u)\frac{\partial\varphi}{\partial x}\right]dxdt$$

$$+\int_{-\infty}^{\infty}v(x)\varphi(x,0)dx=0,\qquad\forall\varphi\in E.$$

这正是(2.39)式，它符合 $u(x,t)$ 是方程(2.35a)满足初始条件 $u(x,0)=v(x)$ 的广义解的定义。

下面再证明(2.44)式的极限函数 $u(x,t)$ 满足熵条件(OE)。为此，在方程(2.44)两端同乘有界函数 $h(u_\varepsilon)$，则得

$$h(u_\varepsilon)\frac{\partial u_\varepsilon}{\partial t}+h(u_\varepsilon)\frac{\partial f(u_\varepsilon)}{\partial x}=\varepsilon h(u_\varepsilon)\frac{\partial^2 u_\varepsilon}{\partial x^2}.\qquad(2.46)$$

令

$$I(u)=\int_{s_0}^{u}h(s)ds,\qquad F(u)=\int_{s_0}^{u}h(s)f'(s)ds,\qquad(2.47)$$

其中 s_0 是任意给定的常数，$I(u)$ 和 $F(u)$ 分别称为熵函数和熵条件，则有

$$\frac{\partial^2 I(u)}{\partial x^2}=\frac{\partial}{\partial x}\left[h(u)\frac{\partial u}{\partial x}\right]=h(u)\frac{\partial^2 u}{\partial x^2}+\frac{\partial h}{\partial u}\left(\frac{\partial u}{\partial x}\right)^2,$$

从而(2.46)式可改写为

$$\frac{\partial I(u_\varepsilon)}{\partial t}+\frac{\partial F(u_\varepsilon)}{\partial x}=\varepsilon\frac{\partial^2 I(u_\varepsilon)}{\partial x^2}-\varepsilon h'(u_\varepsilon)\left(\frac{\partial u_\varepsilon}{\partial x}\right)^2.$$

因此，当 $h(u)$ 无特殊限制时，在上式中令 $\varepsilon\to 0$，则有

$$\frac{\partial I(u_\varepsilon)}{\partial t}+\frac{\partial F(u_\varepsilon)}{\partial x}=0. \tag{2.48}$$

(2.48)式称为**附加守恒律**，它对研究双曲方程的广义解和物理解也是有意义的。

现对 $h(u)$ 加强限制，令 $h'(u)\geqslant 0$，即令 $I(u)$ 是凸函数，则有

$$\frac{\partial I(u_\varepsilon)}{\partial t}+\frac{\partial F(u_\varepsilon)}{\partial x}-\varepsilon\frac{\partial^2 I(u_\varepsilon)}{\partial x^2}\leqslant 0;$$

再令 $\varepsilon\to 0$，则有

$$\frac{\partial I(u)}{\partial t}+\frac{\partial F(u)}{\partial x}\leqslant 0. \tag{2.49}$$

(2.49)称为**附加熵条件**。

设 $D=D_1\cup D_2$，它含有一段间断线 γ，如图2.13所示，对(2.49)式使用 Green 公式，则有在整个间断线上成立的关系式

$$[I(u_+)-I(u_-)]\frac{dx}{dt}\geqslant F(u_+)-F(u_-). \tag{2.50}$$

取

$$s_0=-\infty,\quad h(s)=\begin{cases}0, & \text{当 } s<w \text{ 时},\\ 1, & \text{当 } s\geqslant w \text{ 时},\end{cases}$$

有

$$I(u)=\begin{cases}0, & \text{当 } u<w \text{ 时},\\ u-w, & \text{当 } u\geqslant w \text{ 时},\end{cases}$$

$$F(u)=\begin{cases}0, & \text{当 } u<w \text{ 时},\\ f(u)-f(w), & \text{当 } u\geqslant w \text{ 时}.\end{cases}$$

从而当 $u_-<u_+$ 或 $u_+<u_-$ 时，可从(2.50)式分别推出

$$(u_+-w)\frac{f(u_+)-f(u_-)}{u_+-u_-}\geqslant f(u_+)-f(w),\qquad \forall w\in I,$$

$$(u_--w)\frac{f(u_+)-f(u_-)}{u_+-u_-}\leqslant f(u_-)-f(w),\qquad \forall w\in I,$$

这正是熵条件(2.41a)和(2.41b)式。这就证明了(2.44)式的极限函数 $u(x,t)$ 是满足熵条件(OE)的。 □

前面已就单个守恒型双曲方程 (2.35a) 讨论了广义解、物理解和熵条件的有关问题。这些内容可类似地扩展到守恒型双曲方程组(2.35b)情形中。

例如关于广义解的定义方面，有如下描述。

定义 I 对于在 $t>0$ 上半平面上与 $U(x,t)$ 的间断线只有有限个交点的任意逐段光滑闭回路 Γ，若 $U(x,t)$ 适合积分关系

$$\int_{\Gamma} U\,dx-F(U)\,dt=0, \tag{2.51}$$

则称 $U(x,t)$ 是(2.35b)式的广义解。

定义 II 设 K 表示分片连续可微的函数类，若 $U\in K$ 在解的光滑区域满足微分方程(2.35b)，在间断线上满足关系式

$$(U_+-U_-)\frac{dx}{dt}=F(U_+)-F(U_-), \tag{2.52}$$

则称 $U(x,t)$ 是(2.35b)的广义解。

定义 III 设 $\Phi\in E_\Phi$ 是具有紧致支集的试验函数，若 $U(x,t)$ 满足

$$\int_{t>0}\left[U\frac{\partial\Phi}{\partial t}+F(U)\frac{\partial\Phi}{\partial x}\right]dx\,dt=0,\qquad \forall\Phi\in E_\Phi, \tag{2.53}$$

则称 $U(x,t)$ 是(2.35b)的广义解。

与单个方程的情形不同，(2.35b)式有 n 个不同的实特征值它们可按大小排列为

$$\lambda_1(U)<\lambda_2(U)<\cdots<\lambda_n(U),$$

由 $dx/dt=\lambda_k(U)$, $k=1,2,\cdots,n$ 定义的曲线称为 k-特征线。对于特征线 $x=x(t)$，若不等关系

$$\lambda_{k-1}(U_-)<\frac{dx}{dt}<\lambda_k(U_-)\quad \lambda_k(U_+)<\frac{dx}{dt}<\lambda_{k+1}(U_+)$$

或

$$\lambda_k(U_+)<\frac{dx}{dt}<\lambda_k(U_-) \tag{2.54}$$

成立，这样的间断线 $x=x(t)$ 则称为 k 激波。这个不等式相当于单个方程情形的(CE)熵条件。

例如，Euler 坐标系中的一维流体力学方程组

$$\frac{\partial\rho}{\partial t}+\frac{\partial}{\partial x}(\rho u)=0,$$

$$\frac{\partial(\rho u)}{\partial t}+\frac{\partial}{\partial x}(\rho u^2+p)=0,$$

$$\frac{\partial(\rho E)}{\partial t}+\frac{\partial}{\partial x}(\rho Eu+pu)=0,$$

其中 $E=\left(e+\frac{1}{2}u^2\right)$; e, ρ, u 和 p 分别为内能，密度，速度和压力。这个方程组有三条特征線：

$$\frac{dx}{dt}=u-c,\qquad \frac{dx}{dt}=u,\qquad \frac{dx}{dt}=u+c,$$

其中 c 为声速。方程组的解有三种可能的间断：

(1) $S<u_--c_-$, $u_+-c_+<S<u_+$;

(2) $u_--c_-<S<u_-$, $u_+<S<u_++c_+$;

(3) $u_-<S<u_-+c_-$, $u_++c_+<S$;

其中 S 是间断传播的速度。情况(1)和(3)分别对应于向左和向右的激波。如果在不等式(2.54)中允许等号成立，即间断可以是沿

着特征线传播的，则情况(2)对应于**接触间断**。从(3)可看出

$$S-u_{-}<c_{-},\ S-u_{+}>c_{+},$$

即激波速度相对于波后粒子速度是亚声速的，而相对于波前是超声速的。这一事实等价于过激波的熵是增加的。因此不等式(2.54)也称为熵条件。

对于守恒型双曲方程组(2.35b)也可采用粘性消失法讨论其广义解和物理解。仿(2.44)式有

$$\frac{\partial U_{\varepsilon}}{\partial t}+\frac{\partial F(U_{\varepsilon})}{\partial x}=\varepsilon\frac{\partial^2 U_{\varepsilon}}{\partial x^2},\quad \varepsilon>0. \tag{2.55}$$

不难证明，如 $\varepsilon\to0$ 时，(2.55)式的解 $U_{\varepsilon}(x,t)$一致有界且几乎处处收敛到极限 $U(x,t)$，则 $U(x,t)$一定是(2.35b)式的广义解和物理解。

仿照定理2.2中的推导，用 $H(U)$ 乘(2.55)式两边，并定义**熵函数**和**熵通量**(仿(2.47)式)

$$W(U)=\int_{S_0}^{U}H(s)ds,\quad Q(U)=\int_{S_0}^{U}H(s)F'(s)ds, \tag{2.56}$$

则可得到**附加守恒律**(仿(2.48)式)

$$\frac{\partial W}{\partial t}+\frac{\partial Q}{\partial x}=0. \tag{2.57}$$

若令W是凸函数，则有**附加熵条件**(仿(2.49)式)

$$\frac{\partial W}{\partial t}+\frac{\partial Q}{\partial x}\leqslant0. \tag{2.58}$$

同样可证明下面定理。

定理 2.3 设熵函数和熵通量由(2.56)式定义，则当$W(U)$是凸函数时，(2.55)式的几乎处处有界收敛解 $U(x,t)$，必满足附加守恒律(2.57)式和附加熵条件(2.58)式，且 $U(x,t)$ 是(2.35b)式的物理解。

这个定理表明，人们可利用(2.55)式也可利用(2.57)和

(2.58)式研究双曲方程组(2.35b)的广义解和物理解。有关这方面的深入讨论，读者还可参阅文献[5]~[14]。

6.2 一致差分格式

守恒型双曲方程(组)的解常会发生间断，这给数值求解过程带来不小的困难。一方面，人们要捕捉激波；另一方面，人们还要顾及到激波条件或熵条件。若设守恒型方程(组)的解是分片光滑的，且假设已经捕捉住激波，则数值求解过程可这样进行：在解的光滑部分，使用§1和§2中所介绍的种种差分格式，例如L-F格式、L-W格式、跳蛙格式和特征线方法等等；在间断线上用R-H条件或熵条件联接。但这种方法实现起来有一定困难，主要原因是在一般情况下人们并不知道激波存在于何处。而最关心最感兴趣的是**一致差分格式**，也叫**通用差分格式**，它能在解的光滑部分和间断部分都有统一的形式，且便于计算，它的解是弱解。这就要求人们利用广义解的定义构造出实用的一致差分格式。下面将利用广义解定义Ⅰ和广义解定义Ⅲ特别是定义Ⅲ，构造几种常用的一致差分格式。

一、Lax-Friedrichs格式

因为(2.47)式是积分恒等式，它能适应解的光滑和间断情形，所以由定义Ⅰ构造出的差分格式是能够求出广义解的一致差分格式的。

利用守恒型方程组(2.35b)的积分形式

$$\int_{D_1}\left(\frac{\partial U}{\partial t}+\frac{\partial F}{\partial x}\right)dxdt=\int_{\partial D_1}(Fdt-Udx)=0,$$

取矩形框$ABCD$为∂D_1，如图2.16所示，则有

$$\int_{AB}-Udx+\int_{CD}-Udx+\int_{BC}Fdt+\int_{DA}Fdt$$

$$=\int_{AB}-Udx+\int_{DC}Udx+\int_{BC}Fdt+\int_{AD}-Fdt=0.$$

把其中第一个积分用梯形公式，第二个积分用中矩形公式，后两个积分用下矩形公式去近似，就得到截断误差为 $O(\tau+h^2)$ 的差分格式

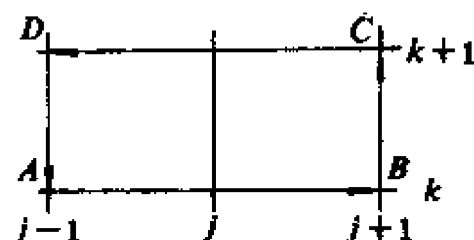


图2.16

$$\frac{1}{\tau}\left[U_j^{k+1}-\frac{1}{2}(U_{j+1}^k+U_{j-1}^k)\right]$$

$$+\frac{1}{2h}(F_{j+1}^k-F_{j-1}^k)=0, \tag{2.59}$$

这就是形如§2.1中的 L-F 格式。它是一个非线性差分格式。为了研究其稳定性，可先把(2.59)式线性化：

$$U_j^{k+1}=\frac{1}{2}(U_{j+1}^k+U_{j-1}^k)-\frac{\tau}{2h}\left[\frac{\partial F}{\partial U}\right]_j^k(U_{j+1}^k-U_{j-1}^k),$$

再使用冻结系数法就得到(2.59)式的稳定性条件是 $\frac{\tau}{h}|\lambda^k|_{max}\leqslant 1$，其中 $|\lambda^k|_{max}=\max\limits_i|\lambda_i^k|$，$\lambda_i^k(i=1,2,\cdots,n)$ 是 Jacobi 矩阵 $[\partial F/\partial U]_j^k$ 的特征值。上述稳定性分析方法称为**线性化稳定性分析法**。

下面利用广义解定义 Ⅲ，具体利用(2.51)式构造一致差分格式。为此，在(2.35b)式中附加粘性项 $\frac{h^2}{2\tau}\frac{\partial^2 U}{\partial x^2}$，就得到拟线性抛物型方程组

$$\frac{\partial U}{\partial t}+\frac{\partial F(U)}{\partial x}=\frac{h^2}{2\tau}\frac{\partial^2 U}{\partial x^2}. \tag{2.60}$$

显然，在网比 $r=\tau/h$ 固定的情况下，当 $h\to0$，$\tau\to0$时，(2.60)式的极限就是(2.35b)式。现将(2.60)式直接差分化离散，得

$$\frac{U_j^{k+1}-U_j^k}{\tau}+\frac{F_{j+1}^k-F_{j-1}^k}{2h}=\frac{h^2}{2\tau}\frac{U_{j+1}^k-2U_j^k+U_{j-1}^k}{h^2}, \tag{2.61}$$

再变形，就得到(2.59)式。这说明了定义Ⅰ和定义Ⅲ的某种统一性。

二 Lax-Wendroff格式

在(2.35b)式中附加粘性项 $\tau\dfrac{\partial}{\partial x}\left(F'(U)\dfrac{\partial F}{\partial x}\right)$，得

$$\frac{\partial U}{\partial t}+\frac{\partial F}{\partial x}=\tau\frac{\partial}{\partial x}\left(F'(U)\frac{\partial F}{\partial x}\right),$$

这是个带小参数 τ 的非线性抛物型方程。当 $\tau\to 0$、$h\to 0$ 时，它的极限也是(2.35b)式。用中心差商将其离散，就有

$$\frac{1}{\tau}(U_j^{n+1}-U_j^n)+\frac{1}{2h}(F_{j+1}^n-F_{j-1}^n)$$

$$-\frac{\tau}{2h^2}[F'(U_{j+1/2}^n)(F_{j+1}^n-F_j^n)$$

$$-F'(U_{j-1/2}^n)(F_j^n-F_{j-1}^n)]=0,$$

整理得

$$U_j^{n+1}-U_j^n+\frac{r}{2}(F_{j+1}^n-F_{j-1}^n)-\frac{r^2}{2}$$

$$\cdot[F'(U_{j+1/2}^n)(F_{j+1}^n-F_j^n)-F'(U_{j-1/2}^n)(F_j^n-F_{j-1}^n)]=0, \tag{2.62}$$

这就是 L-W 格式，它基本与§2.2中的 L-W 格式同形。特别，当 F' 为常数矩阵时，两者完全一样。仿照§1.3中的推导，易知 L-W 格式的截断误差是 $O(\tau^2+h^2)$。同样可用线性化稳定性分析法推出 L-W 格式的稳定性条件是 $\dfrac{\tau}{h}\max\limits_{i}\rho(F'(U_i^n))\leqslant 1$。

三 粘性差分格式

经前面的讨论知道：三个广义解定义中，定义Ⅲ是有代表性的；而且利用这一定义构造了守恒型双曲方程组的 L-F 格式和 L-W 格式。使我们感兴趣的是，它们分别与一阶常(变)系数双

曲方程（组）的 L-F 格式和 L-W 格式同形。这一事实和即将验证的差分格式说明，§1中介绍的种种有效的差分格式都等价于附加不同粘性项后的**粘性差分格式**。例如，§1.2 中的 L-F 格式就等价于在(2.3)式中附加粘性项$\frac{h^2}{\tau}\frac{\partial^2 u}{\partial x^2}$后再离散的差分格式。又如，§1.3中 L-W 格式就等价于在(2.3)式中附加粘性项$a\tau\frac{\partial^2 u}{\partial x^2}$后再离散的差分格式。再如，§1.7中的特征差分格式

$$\frac{u_j^{n+1}-u_j^n}{\tau}+a\frac{u_j^n-u_{j-1}^n}{h}=0$$

可变形为

$$\frac{u_j^{n+1}-u_j^n}{\tau}+a\frac{u_{j+1}^n-u_{j-1}^n}{2h}=\frac{ah}{2}\frac{u_{j+1}^n-2u_j^n+u_{j-1}^n}{h^2},$$

这相当于在模型方程(2.3)中引入粘性项$\frac{ah}{2}\frac{\partial^2 u}{\partial x^2}$后再离散的差分格式，同样也可验证跳蛙格式等相应等价于方程(2.3)中引入不同粘性项的结果。这一事实说明，对模型方程(2.3)所构造的各种差分格式，若它等价于引入不同粘性项的结果，则它一定是能够用于计算线性问题间断解的一致差分格式，而且它一定能够有效地推广到守恒型双曲方程(组)的情形。

附加粘性项是数值求解拟线性双曲型方程(组)中经常使用的手段。人们可以在守恒型方程(组)中先附加粘性项再将其差分化，这样得到的差分格式称为**显式人工粘性格式**；人们也可以在守恒型方程(组)的差分格式中再引入一些相当于粘性项的数值项，或某种差分格式中隐含着一个相当于粘性项的数值项，这样的差分格式称为**隐式人工粘性格式**或**数值粘性格式**，显式和隐式人工粘性格式统称为**粘性差分格式**。粘性差分格式的特点是：不论拟线性双曲型方程(组)有无间断解，可不加区别地按统一格式加以处理。此时，间断是作为解的一部分进行计算的。也就是说

间断解被表示为具有一定过渡层的连续解，这个连续解基本上能反映出间断的特征。

为了观察粘性差分格式处理间断解的特点，现用 L-F 格式和 L-W 格式计算初值问题：

$$\begin{cases} \dfrac{\partial u}{\partial t}+\dfrac{\partial}{\partial x}\left(\dfrac{u^2}{2}\right)=0, \\ u(x,0)=\begin{cases}1, & x\leqslant 0, \\ 0, & x>0,\end{cases}\end{cases}$$

其真解是

$$u(x,t)=\begin{cases}1, & 2x\leqslant t, \\ 0, & 2x>t.\end{cases}$$

这一模型问题是众多学者乐于研究的问题，它能够反映出守恒型(非线性)双曲方程的本质，不同的初值就会产生不同的结果。这里，采用网格步长 $h=\tau=1$，把在 $t=44\tau$ 处的计算结果和真解一起绘置于图2.17中。

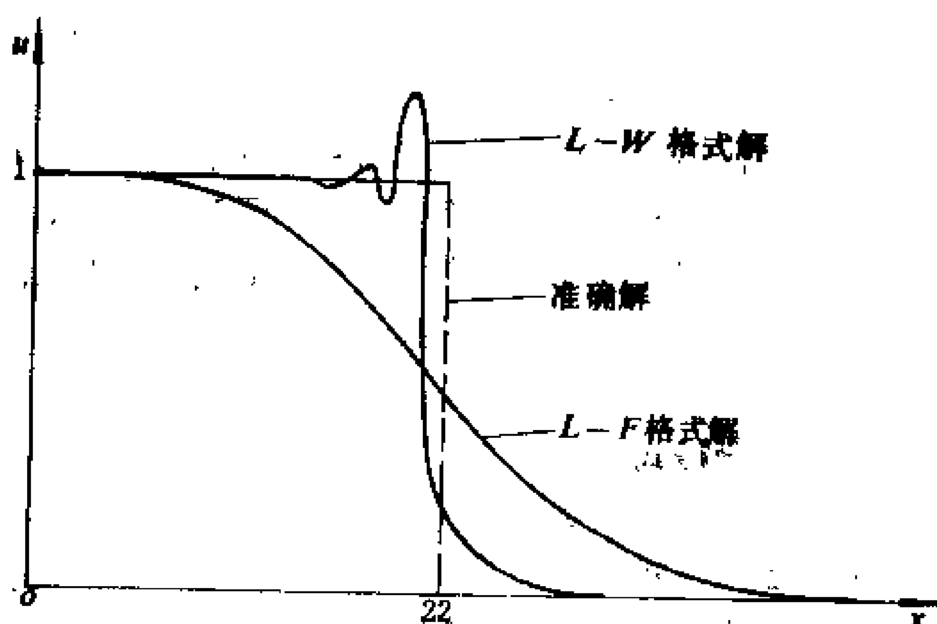


图2.17

从图2.17看出，L-F 格式在解的光滑区逼近效果不够好，

而且它把描述间断的过渡区拉得过宽。这可能是由于L-F格式中引进了粘性过强的粘性项的缘故。L-W格式在解的光滑区的逼近精度较高，描述间断的过渡区也不宽，但在间断线附近出现振荡。这可能是由于L-W格式中引进了粘性过弱且会振荡的粘性项的缘故。

改进粘性格式的逼近效果的办法之一是引入理想的粘性项。这要针对具体方程的特点，使间断部分得到适当磨光，又不使间断强度过于减弱，并能使间断过渡区保持在3～4个网格内。因此合理引入人工粘性项是一项技巧性较强的工作。

改进粘性格式逼近效果的办法之二是组合使用一阶和二阶精度的粘性格式。例如，在解的光滑区使用精度好的L-W格式，在间断线附近使用无振荡的L-F格式。

文献[3]在§6.2.3中就线性问题$\frac{\partial u}{\partial t}+a\frac{\partial u}{\partial x}=f$的情形分析了各种差分格式逼近间断解的效果。文中用局部Fourier分析法对一致差分格式的**粘性效果**和**相变效果**作了分析。粘性效果也就是阻尼效应；分析结果表明：若一致差分格式的精度高，则逼近间断解的效果会好些，而且描述间断解的过渡区也较理想。相变反应了差分解的各种频率函数与真解的各种频率函数之间的相位差；分析表明：相变相对小的一致差分格式要好些，否则就振荡严重。综合来看，相变小的高精度一致差分格式是较理想的。上述结果虽然是对线性情形的分析结果，但它可能对设计守恒型一致差分格式有指导意义。

四　多步差分格式

第一篇§8.3中介绍了解非线性抛物型方程的预测校正格式，这是一种**多步差分格式**。多步差分格式结构简单，计算量小，容易得到高精度效果，所以它在求解守恒型(非线性)双曲方程(组)的算法研究和应用实践中都受到重视。

Richtmyer于1963年将L-W格式(2.62)变形，提出了一

种二步算法：

$$\begin{cases} U_{j+\frac{1}{2}}^{k+\frac{1}{2}} = \frac{1}{2}(U_{j+1}^{k} + U_{j}^{k}) - \frac{r}{2}(F_{j+1}^{k} - F_{j}^{k}) & (2.63a) \\ U_{j}^{k+1} = U_{j}^{k} - r\left(F_{j+\frac{1}{2}}^{k+\frac{1}{2}} - F_{j-\frac{1}{2}}^{k+\frac{1}{2}}\right), & (2.63b) \end{cases}$$

其中$F_{j+\frac{1}{2}}^{k+\frac{1}{2}} = F\left(U_{j+\frac{1}{2}}^{k+\frac{1}{2}}\right)$。第一步(2.63$a$)式是 L-F 格式，第二步(2.63$b$)式是跳蛙格式；第一步可看作预测步，第二步可看作校正步；两步合起来就是二阶精度的 L-W 格式，读者可由线性方程简单验证这一事实。这个二步算法不需要计算 L-W 格式(2.62)中的$F'\left(U_{j+\frac{1}{2}}^{k}\right)$和$F'\left(U_{j-\frac{1}{2}}^{k}\right)$，而且每一个部分步都是显式计算，所以它是非常方便实用的。

二步算法的优点引起了人们的重视，在 Richtmyer 提出 L-W 格式的二步算法后，许多计算流体力学工作者和计算数学工作者提出了不同的二步算法。1973年 Mc Guire 和 Morris 提出了如下的二步格式：

$$\begin{cases} U_{j}^{s} = \frac{1}{2}\left(U_{j+\frac{1}{2}}^{k} + U_{j-\frac{1}{2}}^{k}\right) - sr\left(F_{j+\frac{1}{2}}^{k} - F_{j-\frac{1}{2}}^{k}\right), \\ U_{j}^{k+1} = U_{j}^{k} - \frac{r}{2}\left[\left(1 - \frac{1}{2s}\right)\left((F_{j+1}^{k} - F_{j-1}^{k}\right) \right. \\ \qquad \left. + \frac{1}{s}\left(F_{j+\frac{1}{2}}^{s} - F_{j-\frac{1}{2}}^{s}\right)\right], \end{cases} \tag{2.64}$$

其中$0 < s \leqslant 1$。如果取$s = 1/2$，则(2.64)式化为 Richtmyer 的二步格式。如果取$s = 1$，则(2.64)式就是1967年由 Rubin 和 Burstein 提出的二步格式。可以看出，(2.64)式也是一个显式计算的二步算法。

另一个非常有趣但又与前不同的二步格式是 Mac Cormack

二步格式：

$$\begin{cases}\widetilde{U}_j^k = U_j^k - r(F_{j+1}^k - F_j^k),\\ U_j^{k+1} = \dfrac{1}{2}(U_j^k + \widetilde{U}_j^k) - \dfrac{r}{2}(\widetilde{F}_j^k - \widetilde{F}_{j-1}^k);\end{cases}$$

或

$$\begin{cases}\widetilde{U}_j^k = U_j^k - r(F_j^k - F_{j-1}^k),\\ U_j^{k+1} = \dfrac{1}{2}(U_j^k + \widetilde{U}_j^k) - \dfrac{r}{2}(\widetilde{F}_{j+1}^k - \widetilde{F}_j^k);\end{cases}$$

其中 $F_j^k = F(U_j^k)$，$\widetilde{F}_j^k = F(\widetilde{U}_j^k)$。这个两步算法在两个部分步中交替使用了空间方向的向前差商和向后差商，构造很简单。目前，Mac Cormack 格式已在气体动力学的计算中得到了普遍的应用，并取得了良好的效果。

6.3 一致差分格式的收敛性

我们在§6.2中介绍了几种数值求解守恒型方程组(2.35b)的差分格式，如果其解几乎处处有界收敛，则收敛到某个广义解而不一定是物理解。怎样才能保证一致差分格式的解收敛到物理解呢？为此，需对一致差分格式作统一分析。

定义1 称一致差分格式

$$U_j^{k+1} = U_j^k - \frac{\tau}{h}\left(G_{j+\frac{1}{2}}^k - G_{j-\frac{1}{2}}^k\right), \tag{2.65}$$

其中

$$G_{j+\frac{1}{2}}^k = G(U_{j-l+1}^k,\ U_{j-l+2}^k, \cdots, U_{j+l}^k),\quad j = 0, \pm 1, \cdots,$$

$$G_{j-\frac{1}{2}}^k = G(U_{j-l}^k, U_{j-l+1}^k, \cdots, U_{j+l-1}^k),\quad k = 0, 1, \cdots.$$

是守恒型差分格式。称守恒型差分格式(2.65)是与守恒型方程组(2.35b)**相容**的，如果

$$G(U, U, \cdots, U) = F(U). \tag{2.66}$$

显然，守恒型差分格式(2.65)是对守恒型方程

$$\frac{\partial U}{\partial t}+\frac{\partial G(U)}{\partial x}=0 \tag{2.67}$$

的模拟，这里$G(U)$是 非 线 性 函数；显然，只有当(2.67)式与(2.35b)式相等时，即(2.66)式成立时，守恒型差分 格 式(2.65)才与(2.35b)式守恒型方程相容.

不难验证 L-F 格式(2.59)是与(2.35b)式相容的守恒型差分格式. 事实上，L-F 格式可写成

$$\begin{aligned}U_j^{k+1}&=U_j^k+\frac{1}{2}(U_{j+1}^k-2U_j^k+U_{j-1}^k)-\frac{\tau}{2h}(F_{j+1}^k-F_{j-1}^k)\\&=U_j^k-\frac{\tau}{h}\left(G_{j+\frac{1}{2}}^k-G_{j-\frac{1}{2}}^k\right),\end{aligned}$$

其中

$$G_{j+\frac{1}{2}}^k=\frac{1}{2}(F_j^k+F_{j+1}^k)-\frac{h}{2\tau}(U_{j+1}^k-U_j^k),$$

$$G_{j-\frac{1}{2}}^k=\frac{1}{2}(F_{j-1}^k+F_j^k)-\frac{h}{2\tau}(U_j^k-U_{j-1}^k).$$

同样可验 证 L-W 格 式(2.62)也是与(2.35b)式相容的守恒型差分格式. 此时，只要取

$$\begin{aligned}G_{j+\frac{1}{2}}^k=&\frac{1}{2}(F_{j+1}^k+F_j^k)\\&-\frac{\tau}{2h}F'\left(\frac{U_{j+1}^k+U_j^k}{2}\right)(F_{j+1}^k-F_j^k),\end{aligned}$$

$$\begin{aligned}G_{j-\frac{1}{2}}^k=&\frac{1}{2}(F_j^k+F_{j-1}^k)\\&-\frac{\tau}{2h}F'\left(\frac{U_j^k+U_{j-1}^k}{2}\right)(F_j^k-F_{j-1}^k).\end{aligned}$$

定理2.4 若守恒型差分格式(2.65)与守恒型方程组(2.35b)

相容，且(2.65)式的解几乎处处有界收敛到函数 $U(x,t)$，则 $U(x,t)$ 必是(2.35b)式的一个广义解。

证明 根据守恒型差分格式的定义，把(2.65)式改写成

$$\frac{U_j^{k+1}-U_j^k}{\tau}+\frac{G_{j+\frac{1}{2}}^k-G_{j-\frac{1}{2}}^k}{h}=0,$$

再取试探函数 $\Phi(x,t)\in C_0^\infty(D)$ 的相应离散值 Φ_j^k 乘上式两端且对 j 和 k 求和，就有

$$\sum_{k=0}^{T/\tau}\sum_{j=-\infty}^{+\infty}\Phi_j^k\left[\frac{U_j^{k+1}-U_j^k}{\tau}+\frac{G_{j+\frac{1}{2}}^k-G_{j-\frac{1}{2}}^k}{h}\right]h\tau=0.$$

利用 Abel 分步求和公式(1.51)，上式可写成

$$\sum_{k=0}^{T/\tau}\sum_{j=-\infty}^{+\infty}\frac{\Phi_j^{k-1}-\Phi_j^k}{\tau}U_j^k h\tau+\sum_{k=0}^{T/\tau}\sum_{j=-\infty}^{+\infty}\frac{\Phi_j^k-\Phi_{j+1}^k}{h}G_{j+\frac{1}{2}}^k h\tau=0.$$

当 h、$\tau\to 0$ 时，由于 U_j^k 几乎处处有界收敛到 $U(x,t)$，同时 $G_{j+\frac{1}{2}}^k$ 收敛到 $G(U,U,\cdots,U)$ 即 $F(U)$，故上式趋于积分关系式

$$\int_D\left[\frac{\partial\Phi}{\partial t}U+\frac{\partial\Phi}{\partial x}F(U)\right]dx\,dt=0,$$

所以 $U(x,t)$ 是(2.35b)式的广义解。 □

如前所述，一致差分格式解收敛到相应守恒型双曲方程的广义解，为获得物理解，就必须对一致差分格式附加一些条件。为此，这里介绍最早由 C. K. Gedunov (1959年)提出来的单调差分格式。

已知，双曲型方程的解是初始函数值沿特征线传播的结果，例如一阶常系数双曲方程

$$\frac{\partial u}{\partial t}+a\frac{\partial u}{\partial x}=0$$

满足初始条件 $u(x,0)=v(x)$ 的解为

$$u(x,t)=v(x-at),\qquad t>0.$$

因此，如果初始函数 $v(x)$ 是 x 的单调函数，那么对于任意的 $t>0$，解 $u(x,t)$ 也是 x 的单调函数。于是自然想到，用差分方法解线性双曲型方程初值问题时，如果初始值是单调非增（或非减）的，差分方程的解也应该保持单调非增（或非减）的性质。Gedunov 称这种保持单调性的格式是**单调差分格式**。

G. Jennings 于1974年将 Gedunov 的常系数单调差分格式的概念推广到非线性差分格式上。

定义2 若守恒型差分格式(2.65)的右端项

$$\tilde{G}(U_{j-l}^n,U_{j-l+1}^n,\cdots,U_{j+l}^n)=U_j^n-\frac{\tau}{h}\left(G_{j+\frac{1}{2}}^n-G_{j-\frac{1}{2}}^n\right)$$

关于每个变元都是单调不减的，则称(2.65)式是**守恒型单调格式**。

根据这一定义，只要验证

$$\frac{\partial\tilde{G}}{\partial U_{j+m}^n}\geqslant 0,\qquad m=-l,-l+1,\cdots,l,$$

即知某一差分格式是否为单调差分格式。

例如 L-F 格式(2.59)就是单调守恒型差分格式，因为这时

$$\tilde{G}(U_{j-1}^n,U_j^n,U_{j+1}^n)=\frac{1}{2}(U_{j-1}^n+U_{j+1}^n)$$

$$-\frac{\tau}{2h}[F(U_{j+1}^n)-F(U_{j-1}^n)],$$

$$\frac{\partial\tilde{G}}{\partial U_{j-1}^n}=\frac{1}{2}+\frac{\tau}{2h}F'(U_{j-1}^n),$$

$$\frac{\partial\tilde{G}}{\partial U_{j+1}^n}=\frac{1}{2}-\frac{\tau}{2h}F'(U_{j+1}^n),\quad \frac{\partial\tilde{G}}{\partial U_j^n}=0.$$

当 CFL 条件

$$\frac{\tau}{h}\rho(F'(U_j^n))\leqslant 1$$

成立时，即当格式稳定时，就有

$$\frac{\partial \tilde{G}}{\partial U_{j+m}^n} \geqslant 0, \quad m = -1, 0, 1.$$

因此，L-F 格式是单调守恒型差分格式。

同样可以验证，L-W 格式(2.62)在稳定性条件下不是单调守恒型差分格式。

Hartan, Hyman, Lax 于1976年证明了，一维单个守恒律的单调守恒型差分格式的解如果几乎处处有界收敛，则极限函数是物理解且满足熵条件。

Crandall, Majda 于 1980 年研究了多维单个守恒律的单调守恒型差分格式的解如果几乎处处有界收敛，则极限函数也是物理解且满足熵条件。

利用这一结论不难知道，对于单个守恒型双曲方程的L-F格式是单调守恒型差分格式，它的解收敛到物理解，它的解是与相应的熵条件相容的。人们还验证了单个守恒型双曲方程的 L-W 格式的解不满足熵条件，它不一定收敛到物理解[13]。

对于守恒型双曲方程组的单调差分格式，有如下定理。

定理 2.5 守恒型双曲方程组(2.35b)的单调守恒型差分格式，若其极限函数 $U(x,t)$ 满足附加守恒律(2.57)且 $W(U)$ 是凸函数，则它一定满足熵条件(2.58)。

守恒型单调格式是一阶精度的，这一点从守恒型差分格式的定义式(2.65)中即可看出。那么，是否存在着具有另外一种特性的高精度的守恒型差分格式，它的解是物理解？可惜人们目前对此知道得还很少。总之，在守恒型差分格式的研究方面还存在着很多有待读者去探索的问题。

本章主要介绍了数值求解线性和拟线性双曲型方程(组)的基本思想和方法。在这些数值方法的研究中，完全可以引进数值求

解椭圆型方程和抛物型方程的某些思想和算法，例如§4.4中的Strang可裂格式的思想就可引进到两个自变量的守恒型双曲方程的数值求解中，例如第三篇中介绍的多层网格方法也可引进到双曲型方程的数值求解中。非线性双曲型方程所描述的物理过程要比线性情形复杂得多，关于这方面的理论研究和数值方法研究虽然都在丰富多采地发展着，但仍然远不及对椭圆型和抛物型方程在这些研究方面的深刻程度，还有很多尚不清楚的问题急需人们去研究。当然，人们可以从理论、实验和数值模拟计算等多个方面单独或综合探索研究；但就数值模拟计算而言，构造那些描述运动过程好、数值稳定性好、逼近精度高、收敛速度快的离散格式，是微分方程数值解的根本内容。

习　题

1．为了观察一维热传导方程

$$\begin{cases} \dfrac{\partial u}{\partial t}=\dfrac{\partial^2 u}{\partial x^2}, \quad x\in(-\infty,+\infty),\ t>0, \\ u(x,0)=\varphi(x) \end{cases}$$

的Richardson格式的不稳定性，试写出相应的误差传播方程，在 $r=\tau/h^2=1/10$ 时，作出相应的误差传播图表（参见表1.1）。

2．用古典显格式计算

$$\begin{cases} \dfrac{\partial u}{\partial t}=\dfrac{\partial^2 u}{\partial x^2} \quad (0<x<1,t>0), \\ u(x,0)=x \quad (0\leqslant x\leqslant 1), \\ u(0,t)=t, \quad \dfrac{\partial u}{\partial x}(1,t)=0 \quad (t\geqslant 0), \end{cases}$$

其边界条件如何处理？试写出包括边界条件处理的完整的差分方程组。

3．初边值问题

$$\begin{cases} \dfrac{\partial u}{\partial t} = \dfrac{\partial^2 u}{\partial x^2} & (0<x<1,\ t>0), \\ u(x,0) = \sin\pi x & (0<x<1), \\ u(0,t) = u(1,t) = 0 & (t\geqslant 0) \end{cases}$$

的解析解为

$$u(x,t) = \sin\pi x\cdot\exp(\pi^2 t).$$

试编制古典显格式的计算程序，并打印下述结果：

(1) 取 $h=0.1$，$\tau=0.01$，计算差分解 u_j^k 和误差

$$e_j^k = u_j^k - u(x_j, t_k), \qquad k=1,2,\cdots,10;$$

(2) 取 $h=0.1$，$\tau=0.007$，计算 u_j^k 和 e_j^k，$k=1,2,\cdots,10$;

(3) 取 $h=0.1$，$\tau=0.005$，计算 u_j^k 和 e_j^k，$k=1,2,\cdots,10$.

4. 试用 Fourier 分析法讨论一维热传导方程差分格式

$$\theta\left(\frac{u_j^{k+1}-u_j^{k-1}}{2\tau}\right)+(1-\theta)\left(\frac{u_j^k-u_j^{k-1}}{\tau}\right)=\frac{1}{h^2}\delta_x^2 u_j^k$$

的截断误差、稳定性和差分解的误差阶，其中 $0\leqslant\theta\leqslant 1$ 是实参数，$\frac{1}{h^2}\delta_x^2 u_j^k$ 是 $\frac{\partial^2 u}{\partial x^2}$ 的二阶中心差商近似。

5. 用矩阵分析法分析一维热传导问题的 Richardson 格式、Crank-Nicholson 格式和 Du Fort-Frankel 格式的稳定性。

6. 试用冻结系数法构造逼近

$$\frac{\partial u}{\partial t} = a(x)\frac{\partial^2 u}{\partial x^2} + b(x)\frac{\partial u}{\partial x} + c(x)u$$

的古典隐格式，并用 Fourier 分析法和能量分析法分析该差分格式的稳定性，其中 $a(x)\in C^1$，$\min a(x)>0$，$b(x)$、$c(x)\in C^0$ 且 $\min b(x)>0$，$\min c(x)>0$.

7. 试构造逼近 $\frac{\partial u}{\partial t}=\frac{\partial^2 u}{\partial x^2}+\frac{\partial^2 u}{\partial y^2}$ 的 Du Fort-Frankel 格式，并讨论其相容性、截断误差和稳定性。

8. 试构造抛物型方程

$$\frac{\partial u}{\partial t}=\frac{\partial^2 u}{\partial x^2}+\frac{\partial^2 u}{\partial y^2}+cu\ (c\text{ 是正常数})$$

的古典显格式和C-N格式；讨论其稳定性；并总结在增加cu项前后，其相应的差分格式稳定性的变化规律。

9. 试构造逼近

$$\frac{\partial u}{\partial t}=\frac{\partial^2 u}{\partial x^2}+\frac{\partial^2 u}{\partial y^2}+a\frac{\partial u}{\partial x}+b\frac{\partial u}{\partial y}\quad (a、b\text{为正常数})$$

的6点加权隐格式；分析其稳定性；并总结在增加低阶导数项前后的差分格式稳定性的变化规律。

10. 编制初边值问题

$$\begin{cases}\dfrac{\partial u}{\partial t}=a(x,y)\dfrac{\partial^2 u}{\partial x^2}+b(x,y)\dfrac{\partial^2 u}{\partial y^2}+f(x,y,t), \\ \qquad\qquad (x,y)\in\Omega,\ \ t>0, \\ u(x,y,0)=\varphi(x,y),\quad x\in\Omega, \\ u(0,y,t)=g_1(y,t),\quad u(a,y,t)=g_2(y,t),\quad 0\leqslant t\leqslant T \\ \qquad u(x,0,t)=q_1(x,t),\ u(x,b,t)=q_2(x,t)\end{cases}$$

(其中$\Omega=[0,a]\times[0,b]$)的Crank-Nicholson格式和P-R格式的计算程序。

11. 对于三维模型问题

$$\frac{\partial u}{\partial t}=\frac{\partial^2 u}{\partial x^2}+\frac{\partial^2 u}{\partial y^2}+\frac{\partial^2 u}{\partial z^2},$$ 人们可以将其分解为

$$\frac{1}{3}\ \frac{\partial u}{\partial t}=\frac{\partial^2 u}{\partial x^2},\qquad \frac{1}{3}\ \frac{\partial u}{\partial t}=\frac{\partial^2 u}{\partial y^2},$$

$$\frac{1}{3}\ \frac{\partial u}{\partial t}=\frac{\partial^2 u}{\partial z^2},$$

从而构造出局部一维显格式：

$$
\begin{cases}
u_{l,j,m}^{k+1/3} = \left(1 + \dfrac{\tau}{h^2}\delta_x^2\right) u_{l,j,m}^{k}, \\
u_{l,j,m}^{k+2/3} = \left(1 + \dfrac{\tau}{h^2}\delta_y^2\right) u_{l,j,m}^{k+1/3}, \\
u_{l,j,m}^{k+1} = \left(1 + \dfrac{\tau}{h^2}\delta_z^2\right) u_{l,j,m}^{k+2/3};
\end{cases}
$$

和局部一维 C-N 格式：

$$
\begin{cases}
\left(1 - \dfrac{\tau}{2h^2}\delta_x^2\right) u_{l,j,m}^{k+1/3} = \left(1 + \dfrac{\tau}{2h^2}\delta_x^2\right) u_{l,j,m}^{k}, \\
\left(1 - \dfrac{\tau}{2h^2}\delta_y^2\right) u_{l,j,m}^{k+2/3} = \left(1 + \dfrac{\tau}{2h^2}\delta_y^2\right) u_{l,j,m}^{k+1/3}, \\
\left(1 - \dfrac{\tau}{2h^2}\delta_z^2\right) u_{l,j,m}^{k+1} = \left(1 + \dfrac{\tau}{2h^2}\delta_z^2\right) u_{l,j,m}^{k+2/3}.
\end{cases}
$$

试讨论这两种局部一维差分格式的截断误差和稳定性。

12. 用差分法求解色散方程的初边值问题：

$$
\begin{cases}
\dfrac{\partial u}{\partial t} = a\dfrac{\partial^3 u}{\partial x^3}, & a>0,\ 0<x<1,\ t>0, \\
u(x,0) = f(x), & 0 \leqslant x \leqslant 1, \\
u(0,t) = g_1(t), \quad \dfrac{\partial u}{\partial x}(0,t) = g_2(t), & t>0, \\
u(1,t) = g_3(t), & t>0,
\end{cases}
$$

采用下列数值微分公式：

$$
\left(\frac{\partial u}{\partial t}\right)_j^k = \frac{1}{2}\left[\left(\frac{\partial u}{\partial t}\right)_{j-1/2}^k + \left(\frac{\partial u}{\partial t}\right)_{j+1/2}^k\right] + O(h^2),
$$

$$
\left(\frac{\partial u}{\partial t}\right)_{j\pm 1/2}^k = \frac{1}{\tau}\left[(u)_{j\pm 1/2}^{k+1} - (u)_{j\pm 1/2}^{k}\right] + O(\tau),
$$

$$
\left(\frac{\partial^3 u}{\partial x^3}\right)_j^k = \frac{1}{h^3}\left[(u)_{j+3/2}^k - 3(u)_{j+1/2}^k + 3(u)_{j-1/2}^k - (u)_{j-3/2}^k\right] + O(h^2);
$$

并用 $3(\beta u_{j\pm 1/2}^{k+1} + (1-\beta)u_{j\pm 1/2}^{k})$ 代替上式中的 $3(u)_{j\pm 1/2}^k$；离散边值条件采用

$$\frac{1}{2}(u_{-1/2}^{k}+u_{1/2}^{k})=g_1(k\tau), \qquad \frac{1}{h}(u_{1/2}^{k}-u_{-1/2}^{k})=g_2(k\tau),$$

$$u_j^0=f_j, \quad u_N^k=g_3(k\tau);$$

便可推导出下面的二层差分格式

$$\frac{1}{2\tau}(u_{j+1/2}^{k+1}+u_{j-1/2}^{k+1}-u_{j+1/2}^{k}-u_{j-1/2}^{k})$$

$$=\frac{a}{h^3}[u_{j+3/2}^{k}-3\beta u_{j+1/2}^{k+1}-3(1-\beta)u_{j+1/2}^{k}+3\beta u_{j-1/2}^{k+1}$$

$$+3(1-\beta)u_{j-1/2}^{k}-u_{j-3/2}^{k}].$$

(1) 试写出完整的便于计算的差分方程组;

(2) 叙述显式计算过程;

(3) 分析这个两层格式的数值解精度和稳定性;

(4) 试构造一个三层显式计算的差分格式，且要求这个差分格式是绝对稳定的。

13. 试构造逼近一维对流方程

$$\frac{\partial u}{\partial t}+a\frac{\partial u}{\partial x}=0 \quad (a>0)$$

的 C-N 格式，且讨论其截断误差和稳定性。

14. 讨论解一维对流方程的差分格式

$$\frac{1}{\tau}(u_j^{k+1}-u_j^{k})+\frac{a}{2h}(u_{j+1}^{k+1}-u_{j-1}^{k+1})=0 \quad (a>0)$$

的稳定性条件。

15. 以一维对流方程的差分格式为例，讨论 CFL 条件和 Von-Neumann 条件的关系。

16. 设有解一维对流方程的两个差分格式

$$\begin{cases} u_j^{k+1/2}=\frac{1}{2}(u_{j+1}^{k}+u_{j-1}^{k})-\frac{b}{4}\frac{\tau}{h}(\Delta_x+\nabla_x)u_j^{k}, \\ u_j^{k+1}=u_j^{k}-\frac{b}{2}\frac{\tau}{h}(\Delta_x+\nabla_x)u_j^{k+1/2} \end{cases}$$

和

$$
\begin{cases}
u_{j+1/2}^{k+1/2} = \frac{1}{2}(u_{j+1}^k + u_j^k) - \frac{b}{2} \cdot \frac{\tau}{h}(u_{j+1}^k - u_j^k), \\
u_j^{k+1} = u_j^k - \frac{b\tau}{h}(u_{j+1/2}^{k+1/2} - u_{j-1/2}^{k+1/2}),
\end{cases}
$$

其中 b 是正常数，Δ 和 ∇ 分别表示向前差分和向后差分。

(1) 试将这两个两步差分格式改写为一步差分格式；

(2) 试问这两个差分格式是什么差分格式；

(3) 在每个差分格式中，$k+1/2$ 层和 $k+1$ 层上的数值解精度阶分别是多少？

(4) 试将这两个算法与解常微分方程 $y' = f(x, y)$ 的二阶 Runge-Kutta 算法

$$
\begin{cases}
y_{n+1/2} = y_n + \frac{h}{2} f_n, \quad f_n = f(x_n, y_n), \\
y_{n+1} = y_n + h f_{n+1/2}
\end{cases}
$$

比较，并总结这两个算法的特点。

17. 数值求解

$$
u_t + b u_x = 0 \quad (b > 0)
$$

的 RBM 算法 (Rusanov-Burstein-Mirin, 1970) 如下：

$$
\begin{cases}
u_{j+1/2}^{k*} = \frac{1}{2}(u_{j+1}^k + u_j^k) - \frac{b}{3} \frac{\tau}{h}(u_{j+1}^k - u_{j-1}^k), \\
u_j^{k**} = u_j^k + \frac{2b}{3} \frac{\tau}{h}(u_{j+1/2}^{k*} - u_{j-1/2}^{k*}), \\
u_j^{k+1} = u_j^k - \frac{b}{24} \frac{\tau}{h}(2u_{j-2}^k - 7u_{j-1}^k + 7u_{j+1}^k - 2u_{j+2}^k) \\
\qquad - \frac{3b}{8} \frac{\tau}{h}(u_{j+1}^{k**} - u_{j-1}^{k**}) \\
\qquad - \frac{\omega}{24}(u_{j+2}^k - 4u_{j+1}^k + 6u_j^k - 4u_{j-1}^k + u_{j-2}^k),
\end{cases}
$$

其中 ω 是实参数，且记

$$\mu u_j^k = \frac{1}{2}(u_{j+1/2}^k + u_{j-1/2}^k), \quad \delta u_j^k = (u_{j+1/2}^k - u_{j-1/2}^k).$$

(1) 试将 RBM 算法改写为

$$u_j^{k+1} = u_j^k - \frac{b}{2} \frac{\tau}{h}(\Delta_x + \nabla_x)\left(1 - \frac{1}{6}\delta^2\right)u_j^k$$

$$+\left(\frac{b\tau}{h}\right)^2(\Delta_x - \nabla_x)\left(\frac{1}{2} + \frac{1}{8}\delta^2\right)u_j^k$$

$$-\frac{1}{6}\left(\frac{b\tau}{h}\right)^3(\mu\delta^3)u_j^k - \frac{\omega}{24}(\delta^4)u_j^k;$$

(2) 试将 RBM 算法与三阶 Runge-Kutta 算法比较，并说明 RBM 算法的数值解精度阶为 $O(\tau^3 + h^3)$。

18. 证明实系数二次方程 $\lambda^2 - b\lambda - c = 0$ 的根按模 $\leqslant 1$ 的充要条件是

$$|b| \leqslant 1 - c \leqslant 2.$$

19. 试就二维波动方程构造古典显格式和古典隐格式，并讨论其稳定性条件。

20. 编制数值求解二维波动方程的 L-W 格式和 C-N 格式的计算程序。

21. 对二维一阶双曲型方程组

$$\frac{\partial U}{\partial t} + A\frac{\partial U}{\partial x} + B\frac{\partial U}{\partial y} = 0,$$

其中 A 和 B 均为 $n \times n$ 阶常数矩阵；

(1) 试构造 L-W 格式并指出其截断误差；

(2) 分析其稳定性条件；

(3) 此格式是否是一致差分格式；

(4) 此格式是否是粘性差分格式，若是，则指出此格式所描述的带粘性项的微分方程。

22. 求解双曲型方程

$$\frac{\partial u}{\partial t}+\frac{\partial \varphi}{\partial x}=0$$

的 BB 算法(Boris and Book, 1973)如下:

$$\begin{cases}\tilde{u}_j^{n+1}=u_j^n-\frac{1}{2}\ \frac{\tau}{h}(u_{j+1}^n-u_{j-1}^n)\\ \qquad\qquad +\left[\frac{1}{8}+\frac{1}{2}\left(\frac{\tau}{h}\right)^2\right](u_{j+1}^n-2u_j^n+u_{j-1}^n),\\ u_j^{n+1}=\tilde{u}_j^n-\frac{1}{8}(\tilde{u}_{j+1}^{n+1}-2\tilde{u}_j^{n+1}+\tilde{u}_{j-1}^{n+1}),\end{cases}$$

这个二步算法的稳定性条件是$\left|\frac{\tau}{h}\right|\leqslant\frac{1}{2}$。试分析该算法的二个部分步在引进粘性项方面的特点。

23. 试构造数值求解一阶守恒型方程组$\frac{\partial U}{\partial t}+\frac{\partial \Phi}{\partial x}=0$的二步L-W 算法(参见习题16)和 RBM 算法 (参见习题17)。

24. 数值求解一阶守恒型方程(组)$\frac{\partial U}{\partial t}+\frac{\partial \Phi}{\partial x}=0$的二步 GM 算法(Gourlay and Morris, 1968)如下:

$$\begin{cases}U_j^{n+1/2}=\hat{U}_j^n-a\frac{\tau}{h}(\Delta_x+\nabla_x)\Phi_j^n,\\ U_j^{n+1}=U_j^n-\frac{1}{2}\ \frac{\tau}{h}(\Delta_x+\nabla_x)\left[\left(1-\frac{1}{4a}\right)\Phi_j^n+\frac{1}{4a}\Phi_{j+1}^{n+1/2}\right],\end{cases}$$

其中

$$\hat{U}_j^n=\frac{1}{2}[U_{j+1}^n+U_{j-1}^n],$$

$$(\Delta_x+\nabla_x)U_j^n=U_{j+1}^n-U_{j-1}^n,$$

$\Phi_j^k=\Phi(x_j,t_k)$，a 为正的实参数。

(1) 当$a=\frac{1}{4}$时，二步 GM 算法是否等同于二步 L-W 算法？

(2) 设 $B=\partial\Phi/\partial U$ 是常数矩阵，试证明 GM 算法的传播矩阵是

$$G=I-\frac{1}{2}\left(\frac{\tau}{h}\right)^2 B^2\sin^2\sigma h-i\frac{\tau}{h}B\sin\sigma h$$

$$\cdot\left(1+\frac{1}{4a}(-1+\cos\sigma h)\right),$$

当$\frac{\tau}{h}\leqslant\frac{1}{\sqrt{a}}$时满足Von-Neumann 条件，CFL 条件是$\frac{\tau}{h}\leqslant$ 2，稳定性条件是$a\geqslant\frac{1}{4}$；

(3) 试比较 GM 算法与求解 $y'=f(x,y)$的 Euler 方法

$$\begin{cases}y_{n+1}^*=y_n+hf_n, \\ y_{n+1}=y_n+\omega h(f_n+f_{n+1}^*),\end{cases}\qquad 0<\omega<1$$

关于预测校正过程方面的类同之处。

25．在 Euler 坐标下，一维不定常流动方程为

$$\begin{cases}\frac{\partial\rho}{\partial t}+\frac{\partial\rho u}{\partial x}=0, \\ \frac{\partial u}{\partial t}+u\frac{\partial u}{\partial x}=-\frac{1}{\rho}\frac{\partial p}{\partial x}, \\ \frac{\partial s}{\partial t}+u\frac{\partial s}{\partial x}=0,\end{cases}$$

其中 ρ、p、u、s 分别表示密度、压力、速度和熵，p 是 ρ 和 s 的函数。试将这个方程组化为一阶拟线性方程组

$$\frac{\partial U}{\partial t}+A\frac{\partial U}{\partial x}=0$$

的形式；并构造 L-F 格式和 L-W 格式。

26. 初值问题

$$\begin{cases}\frac{\partial u}{\partial t}+\frac{\partial}{\partial x}\left(\frac{u^2}{2}\right)=0,\\ u(x,0)=\begin{cases}1, & x\leqslant 0,\\ 0, & x>0\end{cases}\end{cases}$$

的解析解是

$$u(x,t)=\begin{cases}1, & \frac{x}{t}\leqslant\frac{1}{2},\\ 0, & \frac{x}{t}>\frac{1}{2}.\end{cases}$$

试构造 L-F 格式和 L-W 格式，并编制程序计算之。

27. 设长度为 L 两端简支的梁发生自由振动，以 $u=u(x,t)$ 表示位移，则 u 为下列混合问题的解：

$$\begin{cases}\frac{\partial^2 u}{\partial t^2}=-a^2\frac{\partial^4 u}{\partial x^4},\\ u(x,0)=\varphi_0(x),\\ \frac{\partial u}{\partial t}(x,0)=\varphi_1(x), \quad 0<x<L,\\ u(0,t)=u(L,t)=0,\\ \frac{\partial^2 u}{\partial x^2}(0,t)=\frac{\partial^2 u}{\partial x^2}(L,t)=0, \quad 0<t\leqslant T.\end{cases}$$

令 $v=\frac{\partial u}{\partial t}$， $w=a\frac{\partial^2 u}{\partial x^2}$，则将梁的振动问题表示为关于 t 的一阶方程组：

$$\begin{cases}\frac{\partial v}{\partial t}=-a\frac{\partial^2 w}{\partial x^2},\\ \frac{\partial w}{\partial t}=a\frac{\partial^2 v}{\partial x^2}.\end{cases}$$

(1) 对方程组的第一个方程采用显式，对第二个方程采用隐式，试构造出这个差分格式；

(2) 试证，此差分格式稳定的充要条件是$\frac{a\tau}{h^2}<\frac{1}{2}$；

(3) 试构造简支梁自由振动问题的绝对稳定的隐格式。

参考文献

[1] R. D. Richtmyer & K. W. Morton, *Difference methods for initial-value problems*, New York-London-Sydney, Interscience Publisher, 1967.

[2] Th. Meis & U. Marcowitz, *Numerische Behandlung partieller Differentialgleichungen*, Springer-Verlag, 1978.

[3] L. Lapidus & G. F. Pinder, *Numerical solution of partial differential equations in science and engineering*, John Wiley & Sons, 1982.

[4] J. Noge ed., *Computational techniques for differential equations*, North-Holland, 1984.

[5] P. D. Lax, *Hyperbolic systems of conservation laws* II, Comm. Pure & Appl. Math., Vol X, 537—566, 1957.

[6] P. D. Lax, *Shock waves and entropy, Report in Courant Institute of Mathematic Sciences* AEC Computing *and* Applied Math. Center, 1971.

[7] P. D. Lax, *Shock waves and entropy*, Proc. *Symposium at the Uni. of Wisconsin, E. H.* Zarantonello ed., 603—634, 1971.

[8] Bjorn Engquist & Stanley Osher, *One sided difference approximations for nonlinear Conservation laws*, Math. Comp. Vol. 36, No. 154, 1981.

[9] Gray Jennings, *Discrete shocks*, Comm. Pure Appl. Math., Vol. XXVII, 25—37, 1974.

[10] R. W. Maccormack & A. J. Paullay, *The influence of the*

computational mesh on accuracy for initial value probles with discontinuous or nonunique solutions, Computers & Fluids, Vol. 2, 339—361, 1974.

[11] A. Y. Le Roux, *Anumerical Conception of Entropy for quasi-linear equations*, Math. of Computation, Vol. 31, No. 140, 848—872, 1977.

[12] Andrew Majda, *Numerical viscosity and the entropy condition*, Comm. Pure & Appl. Math, Vol. XXXII, 797—838, 1979.

[13] A. Harten & P. D. Lax, Comm. Pure Appl. Math. Vol. 29 1976.

[14] J. H. John ed., *Topics in Numer. Anal.* II, Proc. Roy. Irish Acad., Conf. on Numer., 1976.

[15] D. P. Laurie ed., *Numerical solution of partial Differential equations, theory, tools and case studies*, Birkhäuser Verlag, 1982.

第二篇　解椭圆型方程的有限元方法

椭圆型方程描述了定常态物理现象，例如弹性力学中的平衡问题，无粘性流体的无旋运动、亚声速流及渗流问题，位势场（静电磁场和引力场等）问题，热传导中的温度分布，扩散中的浓度分布及导体中的电子密度分布问题等等都可用椭圆型方程的定解问题来描述。尽管线性椭圆型方程的理论研究已较完善[1]，然而对于绝大多数实际问题，因为定解区域和边值条件复杂，解析解极难寻找，所以用各种近似方法对这些实际问题进行数值研究就很有实际意义和发展前景。

虽然用差分方法解椭圆型方程已有不少研究和应用，但60年代初期由中外数学家明确提出的有限元方法却在数值求解各种实际问题方面表现出极大的优越性和生命力，有限元方法的应用研究很快就进入了黄金时代。后来更多的数学家对此方法作了深入研究，终于弄清了有限元方法是逼近论、微分方程和泛函分析等的巧妙结合，并对该方法的理论问题进行了全面透彻地研究。现在，有限元方法已广泛有效地应用于实际问题的数值研究中，已成为解椭圆边值问题的最主要的数值方法。

本篇先以一维问题为例简捷明朗地介绍了有限元方法的思想方法与实现步骤，然后非常自然地将其推广到二维情形，这非常有利于读者很快掌握住有限元方法的本质。在二维问题的讨论中，本篇突出了读者最需要的有关程序设计和提高数值解精度等实用性环节。本篇最后一章集中介绍了有限元方法的有关理论问题，以便读者根据实际情况选用。

第一章　解一维椭圆边值问题的有限元方法

本章以两点边值问题为例，介绍有限元方法的基本思想和步骤，这将是学习第二章内容的基础。

§1　弦平衡问题的两种数学模型

设一根长为l的弦，两端分别固定在点$A(0,\ 0)$和点$B(l,\ 0)$（见图1.1）。当没有外力作用时，它平衡于x轴上；若有一个连续的负荷$f(x)$垂直向下作用在弦上，则弦变形后处于平衡。

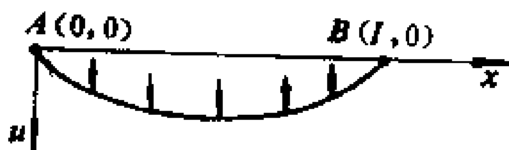


图 1.1

分析弦平衡时的位移$u(x)$有两种途径：一是利用力的平衡关系，知$u(x)$满足微分方程

$$-Tu_{xx}=f(x),\qquad x\in(0,\ l),\tag{1.1}$$

和边值条件

$$u(0)=u(l)=0,\tag{1.2}$$

式中T是弦的张力（设为常数）；二是利用能量守恒定律，即力学中的极小位能原理或虚功原理，也可得到(1.1)和(1.2)式。下面用能量方法讨论弦平衡问题，并讨论它与力分析法之间的联系。

1.1　极小位能原理和变分法

极小位能原理　一个物体在静态平衡时具有极小位能。容易理解，极小位能原理是能量守恒律在静态平衡条件下的应用。这个原理在力学和物理学等学科中有着极其重要的地位，可将其用于分析弦的平衡问题。

弦变形前的能量

$$W_1 = \int_0^l T dx;$$

弦变形后的能量(设弦的变形很小，即$u_{xx} \ll 1$)

$$W_2 = \int_0^l T ds = \int_0^l T(1 + u_x^2)^{1/2} dx \approx \int_0^l T\left(1 + \frac{1}{2}u_x^2\right) dx,$$

弦的应变能（内能）和外力作功分别为

$$W_{内} = W_2 - W_1 \approx \frac{1}{2}\int_0^l T u_x^2 dx, \quad W_{外} = \int_0^l f u dx;$$

因此，据极小位能原理，弦的平衡问题可表述为变分问题：

$$\begin{cases} J(u) = \dfrac{1}{2}\displaystyle\int_0^l T u_x^2 dx - \int_0^l f u dx \Rightarrow \min \\ \forall u \in M: = \{u \in C^2[0, l] \mid u(0) = u(l) = 0\}, \end{cases} \tag{1.3}$$

式中M称为**可供选择的函数类**，又称**容许函数类**；$J(u)$：$M \to R$是函数的函数，称之为**泛函**．(1.3)式是求泛函的极值问题，称为**变分问题**．求解变分问题的方法称为**变分法**[2~4]．1744年，数学家Euler首次提出变分法，并解决了不少在当时被认为神奇且无法解决的变分问题．后经发展，变分法的简捷形式非常容易被人们运用（参见第五章§1)．下面就弦平衡变分问题(1.3)叙述变分法的基本思想．

设$u^*(x)$是(1.3)式的极小解，$\lambda v(x)$是扰动函数，λ是实数，因经扰动后的函数$u^*(x) + \lambda v(x) \in M$，所以对扰动函数$v(x)$的要求是

$$v(x) \in M_{0l} = \{v(x) \in C^2[0, l] \mid v(0) = v(l) = 0\},$$

于是问题(1.3)变为

$$J(u^* + \lambda v) \Rightarrow \min$$

非常有趣且值得特别注意的是，上式是一个关于实参数λ的函数(记为$\varphi(\lambda)$)极值问题，即

$$\varphi(\lambda) = J(u^* + \lambda v) \Rightarrow \min,$$

且在$\lambda=0$时泛函$J(u)$才取得极小值$J(u^*)$。据经典微分法可知，这个泛函取得极小值的必要条件是

$J'_\lambda(u^*+\lambda v)|_{\lambda=0}=0$，即$\varphi'(0)=0$；

为了保证极小解是唯一的，泛函J在极小解处还应满足唯一性条件 $J''_\lambda(u^*+\lambda v)|_{\lambda=0}>0$，即$\varphi''(0)>0$。

这样，利用微分法的知识可求得变分问题的解所满足的条件。

值得指出的是，变分法中设置的扰动函数$v(x)$又称为**试探函数**。虽然试探函数是任意的，但必须满足条件$u^*+\lambda v\in M$。这就是说，试探函数$v(x)$在固定边值点处的函数值等于零，在其它边值点处可任意。具体地，若$u\in M_1=\{u\in C^2[a,b]|u(a)=\alpha, u(b)=\beta\}$，则$v\in M_0:=\{v\in C^2[a,b]|v(a)=v(b)=0\}$；若$u\in M:=\{u\in C^2[a,b]|u(a)=0\}$，则$v\in M_0:=\{v\in C^2[a,b]|v(a)=0\}$。

现在用变分法求弦平衡问题的解所满足的必要条件和唯一性条件。设极小解u^*经扰动后变为$u^*+\lambda v$，u^*_x则变为$u^*_x+\lambda v_x$，那么

$$
\begin{aligned}
\varphi(\lambda)&=J(u^*+\lambda v)\\
&=\frac{1}{2}\int_0^l T(u^*_x+\lambda v_x)^2dx-\int_0^l f(u^*+\lambda v)dx\\
&=\frac{1}{2}\int_0^l T[(u^*_x)^2+2\lambda u^*_x v_x+\lambda^2 v_x^2]dx\\
&\quad-\int_0^l(fu^*+\lambda fv)dx,
\end{aligned}
$$

$$\varphi'(\lambda)|_{\lambda=0}=\int_0^l Tu^*_x v_x dx-\int_0^l fvdx=0, \tag{1.4}$$

$$\varphi''(\lambda)|_{\lambda=0}=\int_0^l Tv_x^2 dx>0. \tag{1.5}$$

若(1.5)式满足，且从(1.4)式中解出u^*，则u^*是(1.3)式的解。(1.4)式又称为**变分方程**，它还可化为一个微分方程。事实上，利用分部积分公式和条件$v(a)=v(b)=0$，(1.4)式则为

$$\int_0^l T u_x^* v_x dx - \int_0^l f v dx$$

$$= \int_0^l T\left(-\frac{d^2u^*}{dx^2}\right) v dx - \int_0^l f v dx + T \cdot \frac{du^*}{dx} v \Big|_0^l$$

$$= \int_0^l \left[-T\frac{d^2u^*}{dx^2} - f\right] v dx = 0, \tag{1.6}$$

由于v的任意性，准确地说，利用泛函分析中的变分预备定理(参见第五章§1) 就有

$$\begin{cases} -T\dfrac{d^2u^*}{dx^2} = f, & x\in(0,\ l), \\ u^*(0) = u^*(l) = 0. \end{cases} \tag{1.7}$$

这就是弦平衡问题的解所满足的微分方程边值问题。由于这是利用Euler变分方法求得的，所以通常称变分问题的解所适合的微分方程为**Euler方程**. 显然，这与力分析法的结果(1.1)和(1.2)式是一致的.

综前所述可以看出，微分方程定解问题和变分问题是可以互相转换的. 对于弦平衡问题，因$u\in C^2$且(1.5)式成立，所以变分问题(1.3)和微分方程边值问题(1.1)及(1.2)都有唯一解，变分问题的解也就是边值问题的解，反之亦然. 有限元方法正是利用这种等价特性，首先将微分方程定解问题转化为变分问题或变分方程的求解问题，然后再设法近似求解变分问题或变分方程.

1.2 虚功原理

要将微分方程在形式上转化为变分方程是非常容易的. 由§1.1中关于Euler方程和变分方程的关系式(1.6)可知，只要将微分方程两端乘上相应的试探函数并积分，即可得到变分方程. 利用这一形式上的特点并加以力学解释，就有以下虚功原理：

虚功原理 若物体所受合力对一切的虚位移（即试探函数）所做的功为零，则物体处于静止平衡态.

因弦所受合力为 $-Td^2u/dx^2-f$，利用虚功原理，弦平衡时必有

$$\int_0^l\left(-T\frac{d^2u}{dx^2}-f\right)vdx=0,\qquad\forall v\in M_0,$$

就有

$$\int_0^l Tu_xv_xdx=\int_0^l fvdx,\qquad\forall v\in M_0,$$

此式称为**虚功方程**，它也就是**变分方程**(1.4)。

为了揭示虚功方程、变分问题和微分方程边值问题三者之间的联系，为了书写方便，我们引进**二次泛函** $a(u,v)$ 和**线性泛函** $F(v)$ 的记号，这里

$$a(u,v):=\int_0^l Tu_xv_xdx,\qquad F(v):=\int_0^l fvdx.$$

于是，变分问题和虚功方程分别有下述形式：

$$J(u)=\frac{1}{2}a(u,u)-F(u)\Rightarrow\min,\qquad\forall u\in M,\tag{1.8}$$

$$a(u,v)=F(v),\qquad\forall v\in M_0;\tag{1.9}$$

(1.5) 式也可相应地改写为

$$a(v,v)>0,\qquad\forall v\in M_0.\tag{1.10}$$

从(1.8)、(1.9)和(1.6)三式，很容易看出微分方程边值问题、变分问题和变分方程在形式上的联系。即先利用虚功原理确定变分方程，再相应改变形式就可构造出相应的变分问题了。

由变分法可知，变分问题(1.8)取得极小解的必要条件和唯一性条件分别是(1.9)和(1.10)式，这表明变分方程的解与变分问题的解（或与微分方程的解）等价的一个重要条件是(1.10)式。这个条件被称为二次泛函 $a(\cdot,\cdot)$ 在试探函数空间 M_0 中的**正定性条件**。因为 $a(\cdot,\cdot)$ 的形式和性质完全由(1.7)式中的微分算子所确定，故又称(1.10)式是微分算子的**椭圆性条件**。

由于二次泛函的正定性（这是由椭圆边值问题有唯一解所确定的）的证明较难，所以将其置于第五章§3中，此外，有兴趣的

读者还可参阅[8]~[15]。为了简明介绍有限元方法的需要，本章只讨论如何建立与微分方程等价的变分形式；只要椭圆型微分方程有唯一解，微分算子的椭圆型条件就自然得到满足了。

1.3 函数类的扩充

在弦平衡变分问题(1.3)和变分方程(1.4)中，需要位移函数$u\in C^2$，这个要求太强了。在实际问题中，常在不连续的载荷力$f(x)$的作用下求弦平衡问题解，此时u是不能属于C^2的；另外从数学上看，弦平衡的变分问题和变分方程只要求u和u_x的平方可以积分就行。据此，完全有必要扩大弦平衡变分形式中的容许函数类，为此，我们引进函数类

$$H^1(a,\ b):=\{u\,|\,u\in L_2(a,\ b);\ u_x\in L_2(a,\ b)\},$$

$$H_0^1(a,\ b):=\{u\,|\,u\in H^1(a,\ b)\text{且}u(a)=u(b)=0\},$$

其中L_2也是一个函数类，它是由一切在$[a,\ b]$上平方可积的函数构成，即

$$L_2(a,\ b):=\left\{f\,\middle|\,\int_a^b|f|^2dx<+\infty\right\}.$$

通常称$H^1(a,\ b)$和$H_0^1(a,\ b)$为**Sobolev空间**，它是由那些函数自身及其一阶导数都平方可积的函数构成的，它要比$C^2(a,\ b)$广大得多。事实上，对H^1中的任一函数u，因其导数u_x是平方可积的，所以u_x可以是阶梯函数，u就可以是分段光滑但整体连续的函数，如折线函数。关于Sobolev空间的简单知识，读者还可参阅第五章§2。

这样，弦平衡变分问题(1.8)可以进一步改述为

$$J(u)=\frac{1}{2}a(u,\ u)-F(u)\Rightarrow\min\quad\forall u\in H_0^1(0,\ l);\tag{1.11}$$

利用变分法，同样可得变分方程

$$a(u,\ v)=F(v),\qquad\forall v\in H_0^1(0,\ l).\tag{1.12}$$

显然(1.11)和(1.12)式比(1.8)和(1.9)式具有更广泛的意义，

(1.11)或(1.12)式的解称为弦平衡问题的**广义解**。

§2 两点边值问题及其等价的变分形式

下面讨论一类重要的两点边值问题

$$
\begin{cases}
Lu=-\dfrac{d}{dx}\left(p(x)\dfrac{du}{dx}\right)+q(x)u=f(x),\, x\in(a,\ b), \\
u(a)=\alpha, \\
p(b)\dfrac{du(b)}{dx}+g(b)u(b)=\beta.
\end{cases} \tag{1.13}
$$

当 $p\in C^1$ 且 $p_{min}>0$；$q\in C^0$ 且 $q\geqslant 0$；$f\in C^0$；$g\in C^0$ 且 $g\geqslant 0$ 时，上式有唯一解 $u\in C^2$。当 $p\in C^0$ 且 $p>0$；q、$f\in L_2$ 且 $q\geqslant 0$，$g\geqslant 0$ 时；上式有唯一解 $u\in H^1$。特别地，当 $p(x)\equiv T$，$q(x)\equiv 0$，$u_x(b)=\alpha=\beta=0$ 时，(1.13)式就是弦平衡问题。形如(1.13)式的两点边值问题在应用和理论研究中是经常出现的[5]。

据§1中讨论的启示，可利用虚功原理先从形式上构造出与(1.13)式相应的变分方程和变分问题。为此先观察：

$$
\begin{aligned}
\int_a^b (Lu-f)v\,dx &= \int_a^b\left[-\frac{d}{dx}\left(p\frac{du}{dx}\right)+qu-f\right]v\,dx \\
&= \int_a^b\left(p\frac{du}{dx}\frac{dv}{dx}+quv-fv\right)dx-p\frac{du}{dx}v\Big|_a^b \\
&= \int_a^b\left(p\frac{du}{dx}\frac{dv}{dx}+quv-fv\right)dx-[\beta-g(b)u(b)]v(b) \\
&= \int_a^b\left(p\frac{du}{dx}\frac{dv}{dx}+quv\right)dx+g(b)u(b)v(b) \\
&\quad -\left[\int_a^b fv\,dx+\beta v(b)\right],
\end{aligned} \tag{1.14}
$$

且定义

$$
a(u,v):=\int_a^b\left(p\frac{du}{dx}\frac{dv}{dx}+quv\right)dx+g(b)u(b)v(b), \tag{1.15}
$$

$$
F(v):=\int_a^b fv\,dx+\beta v(b). \tag{1.16}
$$

再注意(1.14)～(1.16)中的v是试探函数，它是与u在(1.13)式中的边值条件相应的；另外，u在(1.13)式中是二次连续可微的，而在(1.15)、(1.16)式及将来的变分形式中u、v的可微程度也可降低．所以可定义u所属的函数类$H_E^1(a,b)$和试探函数空间$H_{0E}^1(a,b)$如下：

$$u\in H_E^1(a,b):=\{u\in H^1(a,b)\mid u(a)=\alpha\},$$

$$v\in H_{0E}^1(a,b):=\{v\in H^1(a,b)\mid v(a)=0\}.$$

利用§1中关于变分方程和变分问题形式上的特点，我们立即可形式地建立与(1.13)式相应的变分方程和变分问题：

$$a(u,v)=F(v),\quad u\in H_E^1,\ \forall v\in H_{0E}^1,\tag{1.17}$$

$$J(u)=\frac{1}{2}a(u,u)-F(u)\Rightarrow\min\quad \forall u\in H_E^1.\tag{1.18}$$

在证明上述变分形式和微分方程的等价性之前，让我们先观察一下二次泛函$a(u,v)$的几个性质．显然从$a(\cdot,\cdot)$的定义可知它是双线性泛函，即

$$a(u_1+u_2,v)=a(u_1,v)+a(u_2,v),$$

$$a(u,v_1+v_2)=a(u,v_1)+a(u,v_2);$$

也可知$a(u,v)$是**对称**的，即

$$a(u,v)=a(v,u);$$

还可证明$a(v,v)>0$，即$a(v,v)$在试探函数空间中**正定**（证明见第五章§3）．所以称$a(\cdot,\cdot)$是**正定对称的双线性泛函**，这些性质保证了下面两个等价性定理的成立．

定理1.1 若$u\in C^2(a,b)\cap H_E^1(a,b)$，则边值问题(1.13)和变分方程(1.17)等价．

证明 先证必要性．若u是边值问题(1.13)的解且$u\in C^2(a,b)\cap H_E^1(a,b)$，由(1.14)式，有

$$\int_a^b(Lu-f)v\,dx=a(u,v)-F(v)=0,\qquad\forall v\in H_{0E}^1,$$

即u也是变分方程(1.17)的解．

再证充分性. 若u是变分方程(1.17)的解且$u\in C^2(a, b)$,仍由(1.14)式有

$$
\begin{aligned}
&a(u, v)-F(v)\\
&=\int_a^b\left(p\frac{du}{dx}\frac{dv}{dx}+quv\right)dx-\int_a^b fvdx\\
&\quad+[g(b)u(b)-\beta]v(b)\\
&=\int_a^b\left[-\frac{d}{dx}\left(p\frac{du}{dx}\right)+qu\right]vdx+p\frac{du}{dx}v\Big|_a^b-\int_a^b fvdx\\
&\quad+[g(b)u(b)-\beta]v(b)\\
&=\int_a^b (Lu-f)vdx+\left[p(b)\frac{du(b)}{dx}+g(b)u(b)-\beta\right]v(b)=0.
\end{aligned}
$$

由v的任意性(准确地说,由变分预备定理,参见第五章§2.1)即知

$$
Lu-f=0,\qquad x\in(a, b),
$$

$$
p(b)\frac{du(b)}{dx}+g(b)u(b)=\beta.
$$

又因$u\in H_E^1$,所以$u(a)=\alpha$。总之,证得u也是边值问题(1.13)的解. □

定理1.2 若$u\in C^2(a, b)\cap H_E^1(a, b)$,则边值问题(1.13)和变分问题(1.18)等价。

证明 **充分性** 若u^*是变分问题(1.18)的解,对$J(u)$作变分,且利用$a(\cdot,\cdot)$的对称双线性及$F(u)$的线性性质,则对$\forall v\in H_{0E}^1(a, b)$和$\lambda\in R$有

$$
J(u^*+\lambda v)=\frac{1}{2}a(u^*+\lambda v, u^*+\lambda v)-F(u^*+\lambda v)
$$

$$
=J(u^*)+\lambda[a(u^*, v)-F(v)]+\frac{\lambda^2}{2}a(v, v). \qquad (1.19)
$$

因由假设,u^*是变分问题(1.18)的唯一极小解,所以$a(v,v)>0$ $(\forall v\in H_{0E}^1)$,且满足取得极值的必要条件:

$$
J_\lambda'(u^*+\lambda v)\big|_{\lambda=0}=a(u^*, v)-F(v)=0,\qquad \forall v\in H_{0E}^1, \quad (1.20)
$$

即满足变分方程(1.17)。再利用条件$u^*\in C^2\cap H_E^1$和(1.14)式,如同定理1.1的充分性证明,即可证得u^*也是边值问题(1.13)的

解。

必要性 若 u^* 是边值问题（1.13）的解且 $u^*\in C^2(a,\ b)\cap H_E^1(a,b)$，则如同定理1.1中必要性的证明，即知$u^*$必适合变分方程（1.17）。另一方面椭圆型两点边值问题（1.13）有唯一解可导出 $a(v,\ v)>0\ (\forall v\in C^2(a,\ b)\cap H_{0E}^1(a,\ b))$。由这两方面理由且结合（1.19）式即知$u^*$满足变分问题（1.18）取得极小解的必要条件和唯一性条件，所以u^*是变分问题（1.18）的唯一极小解。 □

由定理1.1和定理1.2的证明显见如下事实：

（1）变分问题或变分方程解的光滑性要求比微分方程解的光滑性要求低。与微分方程等价的变分形式的解称作**弱解**，微分方程的解称作**经典解**，只有当$u\in C^2\cap H_E^1$时，弱解和经典解才是一致的。因此，利用变分形式求弱解具有更广泛的实际意义。

（2）对于复杂的边界条件（特别是高维情形），边值问题的求解一般是困难的。若将微分方程转化为相应的变分问题或变分方程，则只需处理第一边值条件，无需处理第二、三边值条件，因为第二、三边值条件已包含于变分形式之中。这一特点对研究微分方程离散化方法及其数值解带来了极大的方便。

（3）利用虚功原理将微分方程化为等价的变分方程（又称虚功方程）是非常方便的。利用这一形式转化的特点，人们可以就各种形式的微分方程（如对称或不对称的、线性或非线性的椭圆型微分方程）形式地构造出与其相应的**弱解方程**（即变分方程）。作为例子，重复本节中构造变分方程的办法，很容易构造出与非对称微分方程边值问题：

$$\begin{cases} Lu=-\dfrac{d}{dx}\left(p(x)\dfrac{du}{dx}\right)+r(x)\dfrac{du}{dx}+q(x)u=f(x), \\ \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad x\in(a,\ b), \\ u(a)=\alpha, \\ p(b)\dfrac{du(b)}{dx}+g(b)u(b)=\beta, \end{cases}$$

相应的变分方程（弱解方程）为

$$a(u,\ v)=F(v),\quad u\in H_E^1(a,\ b),\quad \forall v\in H_{0E}^1(a,\ b),$$

式中$H_E^1(a,\ b):=\{u\in H^1(a,\ b)\,|\,u(a)=\alpha\}$,

$$H_{0E}^1(a,\ b):=\{v\in H^1(a,\ b)\,|\,v(a)=0\},$$

$$a(u,\ v):=\int_a^b\left[p(x)\frac{du}{dx}\frac{dv}{dx}+r(x)\frac{du}{dx}v+q(x)uv\right]dx+g(b)u(b)v(b),$$

$$F(v):=\int_a^b fv\,dx+\beta v(b).$$

§3 Ritz-Galerkin方法

一切静态稳定的物理问题，都可用极小位能原理建立相应的变分问题，但求变分问题的解却不是一件容易的事。在漫长的一段时间里，人们总是利用变分法把变分问题转化为一个相应的Euler方程（即微分方程）求解。但这带来两个困难。其一，解微分方程得到的经典解并不一定是变分问题要求的弱解，这样的经典解有时与实际问题不符合。这是因为将变分问题求解转化为微分方程求解后，提高了对解的光滑性要求。其二，对复杂的区域（如高维情形）和复杂的边界条件解微分方程更是困难，甚至找不出解的表达式。这些困难都要求人们研究变分问题或弱解方程近似解的办法。

1909年，W. Ritz在“求解数学物理变分问题的一种新方法”一文中提出了变分问题的**直接解法**，找到了求变分问题近似解的新路子。后来Galerkin将Ritz方法用于求解虚功方程，使直接解法的应用范围得以扩大。Ritz-Galerkin方法的基本思想是用有限维允许函数直接代替变分问题（或变分方程）中的无限维允许函数，从而在有限维函数空间中求变分问题（或变分方程）的近似解[17]，并且要求当有限维容许函数空间的维数不断

增加时，有限维近似解逼近原变分问题(或变分方程)的解。

Ritz-Galerkin方法求解变分问题(1.18)（或变分方程(1.17)）的基本过程解释如下：

1. 选取一组完全的基函数$\varphi_i(x)$ $(i=1,2,\cdots)$和一个满足第一边值条件的函数$u_0(x)$，使得

$$u_0(a)=\alpha,\ \varphi_i(a)=0,\quad i=1,\ 2,\ \cdots.$$

这样$H^1_{0E}:=\text{span}\{\varphi_1(x),\ \varphi_2(x),\ \cdots\}$,

$$v(x)=\sum_{j=1}^{\infty}d_j\varphi_j(x),\quad \forall v(x)\in H^1_{0E};$$

$$H^1_E:=\text{span}\{u_0(x),\ \varphi_1(x),\ \varphi_2(x),\ \cdots\},$$

$$u(x)=u_0(x)+\sum_{i=1}^{\infty}c_i\varphi_i(x);$$

其中c_i和d_j为自由选择的参数。

2. 取n个基函数构成n维容许函数类V_n和n维试探函数空间V_n^0，即

$$V_n:=\text{span}\{u_0(x),\ \varphi_1(x),\ \varphi_2(x),\ \cdots,\ \varphi_n(x)\},$$

$$u_n=u_0(x)+\sum_{i=1}^{n}c_i\varphi_i(x),\quad \forall u_n\in V_n;$$

$$V_n^0:=\text{span}\{\varphi_1(x),\ \varphi_2(x),\ \cdots,\ \varphi_n(x)\},$$

$$v_n=\sum_{j=1}^{n}d_j\varphi_j(x).$$

3. 在有限维容许函数类V_n中（而不是在无穷维的H^1_E中）寻求变分问题（或变分方程)的近似解。为此，将u_n和v_n的表达式代入变分问题的简写形式(1.18)式，就得到一个在有限维容许函数类V_n中的变分问题：

$$J(u_n)=\frac{1}{2}a(u_n,\ u_n)-F(u_n)\Rightarrow\min\quad \forall u_n\in V_n,$$

或即

$$J(u_n)=\frac{1}{2}a\Big(u_0+\sum_{i=1}^{n}c_i\varphi_i,$$

$$u_0+\sum_{i=1}^{n}c_i\varphi_i\Big)-F\Big(u_0+\sum_{i=1}^{n}c_i\varphi_i\Big)$$

$$=\Big[\frac{1}{2}a(u_0,\ u_0)-F(u_0)\Big]+a\Big(u_0,\ \sum_{i=1}^{n}c_i\varphi_i\Big)$$

$$+\frac{1}{2}a\Big(\sum_{i=1}^{n}c_i\varphi_i,\ \sum_{i=1}^{n}c_i\varphi_i\Big)-F\Big(\sum_{i=1}^{n}c_i\varphi_i\Big)$$

$$=\text{const.}+a\Big(u_0,\ \sum_{i=1}^{n}c_i\varphi_i\Big)$$

$$+\frac{1}{2}a\Big(\sum_{i=1}^{n}c_i\varphi_i,\ \sum_{i=1}^{n}c_i\varphi_i\Big)-F\Big(\sum_{i=1}^{n}c_i\varphi_i\Big)\Rightarrow\min.$$

注意，这是关于 n 个变元 c_1，c_2，…，c_n 的二次函数的极值问题，利用经典微分学中关于多元函数求极值的办法即得到函数 $J(u_n)$ 取得极小值的必要条件：

$$\partial J(u_n)/\partial c_j=a(u_0,\ \varphi_j)+\frac{1}{2}a\Big(\varphi_j,\ \sum_{i=1}^{n}c_i\varphi_i\Big)$$

$$+\frac{1}{2}a\Big(\sum_{i=1}^{n}c_i\varphi_i,\ \varphi_j\Big)-F(\varphi_j)$$

$$=a\Big(u_0+\sum_{i=1}^{n}c_i\varphi_i,\ \varphi_j\Big)-F(\varphi_j)=0,\quad j=1,\ 2,\ \cdots,\ n.$$

这完全等同于将 u_n 和 v_n 表达式代入变分方程(1.17)式的结果。事实上

$$a(u_n,\ v_n)=F(v_n),\qquad \forall v_n\in V_n^0,$$

可写成

$$a\Big(u_0+\sum_{i=1}^{n}c_i\varphi_i,\ \sum_{j=1}^{n}d_j\varphi_j\Big)=F\Big(\sum_{j=1}^{n}d_j\varphi_j\Big),$$

$$\sum_{j=1}^{n}d_j a\Big(u_0+\sum_{i=1}^{n}c_i\varphi_i,\ \varphi_j\Big)=\sum_{j=1}^{n}d_jF(\varphi_j).$$

利用 v_n 的任意性，不妨取 $d_j=(0,\ \cdots,\ 0,\ 1,\ 0,\ \cdots,\ 0)$，即 d_j 为 n 维数组且仅在第 j 个位置上等于 1，其余分量全为零，那么上式就可写成

$$a\Big(u_0+\sum_{i=1}^{n}c_i\varphi_i,\ \varphi_j\Big)=F(\varphi_j),\qquad j=1,\ 2,\ \cdots,\ n.$$

的形式。因此当$a(\cdot, \cdot)$对称正定时，无论是Ritz法还是Galerkin方法，都是在有限维函数空间中求变分问题（或变分方程）的近似解，并最后归结为同一个线性代数方程组（称为Ritz-Galerkin方程组）：

$$\sum_{i=1}^{n} a(\varphi_i, \varphi_j)c_i = F(\varphi_j) - a(u_0, \varphi_j), \quad j = 1, 2, \cdots, n,$$

即

$$\begin{pmatrix} a(\varphi_1, \varphi_1) \cdots a(\varphi_n, \varphi_1) \\ \cdots \quad \cdots \quad \cdots \\ a(\varphi_1, \varphi_n) \cdots a(\varphi_n, \varphi_n) \end{pmatrix} \begin{pmatrix} c_1 \\ \vdots \\ c_n \end{pmatrix} = \begin{pmatrix} F(\varphi_1) - a(u_0, \varphi_1) \\ \cdots \\ F(\varphi_n) - a(u_0, \varphi_n) \end{pmatrix}, \tag{1.21}$$

的求解问题。

由二次泛函$a(\cdot, \cdot)$的正定对称性决定了(1.21)式系数矩阵也是正定对称的，因此(1.21)式可解。解出$c_1, c_2, \cdots, c_n$后即可求得(1.17)或(1.18)式的近似解$u_n = u_0(x) + \sum_{i=1}^{n} c_i\varphi_i(x)$，$x \in [a, b]$，且当$n \to \infty$时$u_n \to u$。

尽管Ritz法和Galerkin法导出的近似解u_n及计算方法完全一样，但二者的基础不同。Ritz法基于极小位能原理，而Galerkin法基于虚功原理，所以Galerkin法较Ritz法应用更广，方法推导也更直接。仅当$a(u, v)$正定对称时，两者才一致；否则只能用Galerkin法而不能用Ritz法。

总上可见，Ritz-Galerkin方法也完全适用于直接近似求解高维问题；另外，Ritz-Galerkin近似解还具有许多良好的数学性质(参见第五章§4)，因此Ritz-Galerkin方法是简明而优美的。

传统的Ritz-Galerkin方法通常选用代数多项式或三角多项式或正交多项式作基函数，为满足零边值条件，还要在此基础上乘一个因式。例如当$\Omega = (0, 1)$且基函数在区间两个端点都为零时，可选用$x(1-x)x^{i-1}$ $(i = 1, 2, \cdots)$或$\sin(i\pi x)$ $(i = 1, 2, \cdots)$等；对$\Omega = (0, 1)^2$且当基函数在$\partial\Omega$上取零值时，可选取

$x(1-x)y(1-y)x^{k-1}y^{l-1}$或$\sin(k\pi x)\sin(l\pi y)(k,l=1,2,\cdots)$等，对$\Omega=\{(r\cos\varphi,\ r\sin\varphi)\mid r\in[0,1),\ \varphi\in[-\pi,\pi)\}$，可选用$r^2-1$、$(r^2-1)\sin\varphi$、$(r^2-1)\cos\varphi$、$(r^2-1)\sin2\varphi$、$(r^2-1)\cos2\varphi$，…作基函数。

作为例子，用传统的Ritz-Galerkin方法解边值问题：

$$\begin{cases}-u''+u=-x, & x\in(0,\ 1);\\ u(0)=u(1)=0.\end{cases} \tag{1.22}$$

选择完全的且满足齐次边值条件的基函数

$$\varphi_i(x)=x(1-x)x^{i-1},\quad i=1,\ 2,\cdots,$$

从而任一有限维允许函数$u_n(x)$可写成

$$u_n(x)=\sum_{i=1}^{n}c_i\varphi_i(x)=x(1-x)(c_1+c_2x+\cdots+c_nx^{n-1}).$$

先令$n=1$，则$u_1=c_1x(1-x)$。此时，Ritz-Galerkin方程为

$$c_1\int_0^1[(\varphi_1')^2+(\varphi_1)^2]dx=\int_0^1-x^2(1-x)dx,$$

解得$c_1=\frac{5}{18}$，从而得$u_1=\frac{5}{18}x(1-x)$。

再令$n=2$，则$u_2=c_1x(1-x)+c_2x^2(1-x)$。经简单计算得Ritz-Galerkin方程为

$$\begin{cases}-\frac{3}{10}c_1-\frac{3}{20}c_2=-\frac{1}{12},\\ -\frac{3}{20}c_1-\frac{13}{105}c_2=-\frac{1}{20}.\end{cases}$$

解之，得$c_1=\frac{71}{369}$，$c_2=\frac{7}{41}$，于是$u_2=x(1-x)\left(\frac{71}{369}+\frac{7}{41}x\right)$。

边值问题(1.22)的精确解是$u(x)=\sin x/\sin1-x$。下表列出了精确解$u(x)$和近似解$u_1(x)$与$u_2(x)$在$x=1/4$，$1/2$，$3/4$处的函数值。从表中数值可见，利用Ritz-Galerkin方法获得的近

似解具有良好的精度.

x	$u(x)$	$u_1(x)$	$u_2(x)$
1/4	0.044	0.052	0.044
1/2	0.070	0.069	0.069
3/4	0.060	0.052	0.060

但是，实际应用中的问题要比上面的例子复杂得多，按传统取基函数的方法用Ritz-Galerkin方法求解会遇到许多原则性的困难。一方面，对复杂区域（如高维曲边界区域），要构造出在整个区域上光滑且满足边界条件的一组完全的基函数是极其困难甚至是不可能的；另一方面，形成Ritz-Galerkin方程时需要惊人的计算量和存贮量，例如，对于取n个基函数的Ritz-Galerkin方程组且考虑到对称性，就要计算并存贮$n(n+3)/2$个积分；此外，还有求解Ritz-Galerkin方程方面的困难，按照传统基函数的取法，方程组的条件数大，会出现数值不稳定，无论用迭代法或消去法都会遇到很大困难。由于上述困难和计算工具不理想等原因，虽然1909年就已提出这种优美的直接解法，但无法用于实际计算，Ritz-Galerkin方法长期不被人们所重视。

60年代初，我国应用数学家冯康和国外的应用数学家几乎同时提出了**有限元方法**，该方法成功地保持了Ritz-Galerkin方法的优点，并克服其缺点使直接法又得新生。有限元法之所以被人们重视和广泛被应用的原因在于：(1) 充分发挥了变分形式和Ritz-Galerkin 方法的优点；(2) 摆脱了传统的基函数取法，方便地构造出局部非零的分段(片)多项式有限维试探函数空间的基函数，从而成功地克服了传统Ritz-Galerkin方法在应用中的缺点；(3) 各种问题的计算程序格式统一，便于计算机自动处理。这些特点使有限元方法获得了巨大的成就，请读者在学习和应用有限元方法的过程中始终注意体会这三个方面。

§4 有限元方法及其步骤

有限元方法首先将微分方程化为一个等价的变分方程(或变分问题)，再用Ritz-Galerkin方法在有限维函数空间中求变分方程的近似解。有限元方法与传统Ritz-Galerkin方法的差别在于有限维函数空间的构造方面。传统的Ritz-Galerkin方法所选用的基函数在整个定解区域上整体光滑，而有限元法则取分段(片)连续且局部非零的基函数。这种差别是关键的，它不仅使有限元法在结构与做法上产生了许多固有的特点，而且使有限元法取得了传统Ritz-Galerkin方法所不能比拟的巨大成就。下面将叙述有限元方法的几个主要步骤。

4.1 区域分划

用节点 $a=x_0<x_1<\cdots<x_{n-1}<x_n=b$ 将区间$[a, b]$分划为 n 个不重迭的小区间（称作**单元**）$E_{i-1}=[x_{i-1}, x_i]$，$i=1, 2, \cdots, n$；记$h_i=x_{i+1}-x_i$，$h=\max h_i$，要求当$n\to\infty$时$h\to 0$。一般情况下可采用均匀分划(即$h_i=h$，$i=1, \cdots, n$)，但特殊情况下可据实际问题的需要，采用非均匀分划也是有益的。例如，在函数值变化较大的地方单元可分得细密些，在别处单元可分得稀疏些。

4.2 线性元

为了对变分形式使用Ritz-Galerkin方法，关键是选取基函数，构造有限维试探函数空间$V_h^*\subset H_{0E}^1$和有限维容许函数类$V_h\subset H_E^1$，且使Ritz-Galerkin方程组简化易解。显然，分段线性多项式函数空间是可以适合要求的，这类函数的特点如图1.2所示。

下面将确定这类函数的基函数。因为在某个单元E(不妨设为$[x_{i-1}, x_i]$)上$u_{h,E}(x)$是线性函数，所以

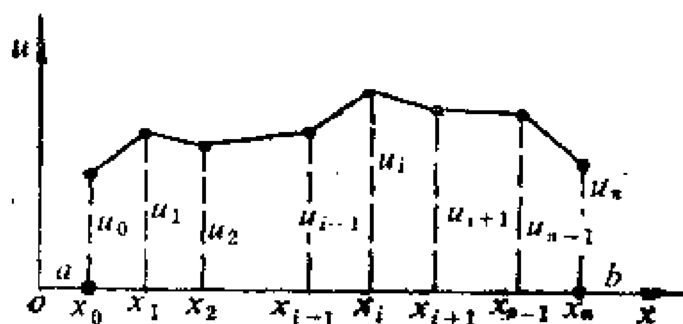


图 1.2

$$u_{h,E}(x)=u_{i-1}\frac{x_i-x}{h_{i-1}}+u_i\frac{x-x_{i-1}}{h_{i-1}},\qquad x\in E,$$

式中，$(x_i-x)/h_{i-1}$ 和 $(x-x_{i-1})/h_{i-1}$ 分别是对应于单元节点 x_{i-1}和x_i处的**单元节点基函数**。若在每个单元上都如此作出线性多项式表达式，则可得到在$[a,\ b]$区间上的分段线性多项式函数$u_h(x)$，且可表示为线性组合的形式：

$u_h(x)=\sum\limits_{i=0}^{n}u_i\varphi_i(x),\quad x\in[a,\ b]$，式中$\varphi_i(x)$称为**整体节点基函数**，其表达式为

$$\begin{cases}\varphi_i(x)=\begin{cases}(x-x_{i-1})/h_{i-1}, & x\in[x_{i-1},\ x_i],\\ (x_{i+1}-x)/h_i, & x\in[x_i,\ x_{i+1}],\\ 0, & \text{别处},\ i=1,\ \cdots,\ n-1,\end{cases}\\ \varphi_0(x)=\begin{cases}(x_1-x)/h_0, & x\in[x_0,\ x_1],\\ 0, & \text{别处},\end{cases}\\ \varphi_n(x)=\begin{cases}(x-x_{n-1})/h_{n-1}, & x\in[x_{n-1},\ x_n],\\ 0, & \text{别处}.\end{cases}\end{cases}\tag{1.23}$$

$\varphi_i(x)$的形状如图1.3所示，$\varphi_i(x)$也称作**山形函数**，它有如下特点：

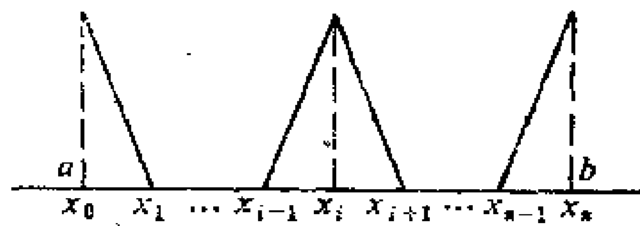


图 1.3

$$
\begin{cases}
\varphi_i(x_j) = \delta_{ij} = \begin{cases} 1, & i = j, \\ 0, & i \neq j, \end{cases} \\
\varphi_i(x) \cdot \varphi_j(x) = 0, \quad |i-j| \geqslant 2, \\
\varphi_i'(x) \cdot \varphi_j'(x) = 0, \quad |i-j| \geqslant 2.
\end{cases} \tag{1.24}
$$

用这类线性插值型基函数，可以构造出适合各种边值条件的**有限维试探函数空间**V_h^0和容许函数类V_h。例如方程的边值条件是$u(a)=\alpha$，$u(b)=\beta$，那么相应地有

$$V_h = \text{span}\{\varphi_0, \varphi_1, \cdots, \varphi_n\}, \quad V_h^0 = \text{span}\{\varphi_1, \cdots, \varphi_{n-1}\};$$

如方程边值条件是$u(a)=\alpha$，$u'(b)=\beta$，那么相应地有

$$V_h = \text{span}\{\varphi_0, \varphi_1, \cdots, \varphi_n\}, \quad V_h^0 = \text{span}\{\varphi_1, \varphi_2, \cdots, \varphi_n\}.$$

4.3 有限元方程的形成和计算

考虑两点边值问题(1.13)，将$\varphi_i(x)$代入相应的Ritz-Galerkin方程(1.21)，就得到有限元方程组

$$KU = B, \tag{1.25}$$

其中

$$K = \begin{pmatrix} a(\varphi_1, \varphi_1) \cdots a(\varphi_n, \varphi_1) \\ \cdots \quad \cdots \\ a(\varphi_1, \varphi_n) \cdots a(\varphi_n, \varphi_n) \end{pmatrix}, \quad U = \begin{pmatrix} u_1 \\ \vdots \\ u_n \end{pmatrix},$$

$$B = \begin{pmatrix} F(\varphi_1) - a(\alpha\varphi_0, \varphi_1) \\ \cdots \\ F(\varphi_n) - a(\alpha\varphi_0, \varphi_n) \end{pmatrix}.$$

因为K和B在力学上反映了物体的刚度和所受的载荷，所以又称K为**刚度矩阵**，B为**载荷向量**，由(1.24)式知，K是3对角矩阵。

用有限元法解与边值问题(1.13)相应的弱解方程时，K和B中元素的计算公式如下($x = x_i + \xi h_i$，$i = 1, 2, \cdots, n-1$)：

$$a(\varphi_i, \varphi_i) = \int_{x_{i-1}}^{x_i} \cdot + \int_{x_i}^{x_{i+1}} \cdot$$

$$= h_{i-1}^{-1} \int_0^1 p(x_{i-1} + \xi h_{i-1}) d\xi + h_{i-1} \int_0^1 q(x_i + \xi h_{i-1}) \xi^2 d\xi$$

$$+h_i^{-1}\int_0^1 p(x_i+\xi h_i)d\xi+h_i\int_0^1 q(x_i+\xi h_{i-1})(1-\xi)^2 d\xi,$$

$$i=1,\ 2,\ \cdots,\ n-1, \quad (1.26)$$

$$a(\varphi_i,\ \varphi_{i+1})=\int_{x_i}^{x_{i+1}}\cdot=-h_i^{-1}\int_0^1 p(x_i+\xi h_i)d\xi$$

$$+h_i\int_0^1 q(x_i+\xi h_i)\xi(1-\xi)d\xi,\quad i=0,1,\cdots,n-1, \quad (1.27)$$

$$a(\varphi_{i-1},\varphi_i)=\int_{x_{i-1}}^{x_i}\cdot=-h_{i-1}^{-1}\int_0^1 p(x_{i-1}+\xi h_{i-1})d\xi$$

$$+h_{i-1}\int_0^1 q(x_{i-1}+\xi h_{i-1})\xi(1-\xi)d\xi,$$

$$i=1,\ 2,\ \cdots,\ n-1, \quad (1.28)$$

$$a(\varphi_n,\ \varphi_n)=\int_{x_{n-1}}^{x_n}\cdot+g(b)[\varphi_n(b)]^2$$

$$=h_{n-1}^{-1}\int_0^1 p(x_{n-1}+\xi h_{n-1})d\xi$$

$$+h_{n-1}\int_0^1 q(x_{n-1}+\xi h_{n-1})\xi^2 d\xi+g(b) \quad (1.29)$$

$$a(u_0,\ \varphi_j)=a(\alpha\varphi_0,\ \varphi_j)=\begin{cases}\alpha a(\varphi_0,\ \varphi_1), & j=1,\\ 0, & j\geqslant 1,\end{cases} \quad (1.30)$$

$$F(\varphi_j)=\int_a^b f(x)\varphi_j(x)dx+\beta\varphi_j(b)=\int_a^b f(x)\varphi_j(x)dx$$

$$=h_{j-1}\int_0^1 f(x_{j-1}+\xi h_{j-1})\xi d\xi+h_j\int_0^1 f(x_j+\xi h_j)(1-\xi)d\xi,$$

$$j=1,\ 2,\ \cdots,\ n-1, \quad (1.31)$$

$$F(\varphi_n)=\int_a^b f(x)\varphi_n(x)dx+\beta\varphi_n(b)=\int_{x_{n-1}}^{x_n}\cdot+\beta$$

$$=h_{n-1}\int_0^1 f(x_{n-1}+\xi h_{n-1})(1-\xi)d\xi+\beta, \quad (1.32)$$

其中$\int_{x_{i-1}}^{x_i}\cdot$表示在单元$[x_{i-1},\ x_i]$上相应的积分。

特别地，对于边值问题

$$
\begin{cases}
-\dfrac{d^2u}{dx^2}=f(x), \quad x\in(a,\ b); \\
u(a)=\alpha, \\
\dfrac{du(b)}{dx}+g(b)u(b)=\beta,
\end{cases}
\tag{1.33}
$$

其有限方程中的诸积分很容易算出，利用(1.26)～(1.28)式和(1.31)式，可写出有限元方程组中第i行($i=1,\ 2,\ \cdots,\ n-1$)的元素为

$$
\begin{cases}
a(\varphi_i,\ \varphi_i)=\dfrac{1}{h_{i-1}}+\dfrac{1}{h_i}, \\
a(\varphi_i,\varphi_{i+1})=-\dfrac{1}{h_i}, \\
a(\varphi_i,\varphi_{i-1})=-\dfrac{1}{h_{i-1}}, \\
F(\varphi_i)=\dfrac{1}{2}(h_{i-1}+h_i)f(x_i).
\end{cases}
\tag{1.34}
$$

若采用等距分划，$h_i\equiv h$，则与(1.25)式相应的有限元方程组为

$$
\begin{pmatrix}
2 & -1 & & & \\
-1 & 2 & -1 & & \\
 & & \ddots & & \\
 & & -1 & 2 & -1 \\
 & & & -1 & 1+g(b)
\end{pmatrix}
\begin{pmatrix}
u_1 \\ u_2 \\ \vdots \\ u_{n-1} \\ u_n
\end{pmatrix}
=
\begin{pmatrix}
\dfrac{1}{2}h^2(f_0+f_1)-\alpha \\
\dfrac{1}{2}h^2(f_1+f_2) \\
\cdots \\
\dfrac{1}{2}h^2(f_{n-2}+f_{n-1}) \\
\dfrac{1}{2}h^2f_{n-1}+\beta
\end{pmatrix}.
$$

一般情形，$p(x)$、$q(x)$和$f(x)$都是函数，那么可编制子程序且用Gauss求积公式计算。但采用多高精度的求积公式与有限元解所能达到的精度密切相关。当然，使用过高精度的求积公式是没有必要的，但也决不应当由于采用了低精度的求积公式以致损害了有限元解可以达到的精度。

已经证明，对于线性元，采用中点Gauss求积公式就可以了。设单元 E_i 的中点 $\bar{x}_i=(x_i+x_{i+1})/2$，则可用下列具有二阶精度的中点求积公式：

$$\int_{E_i} p(x)dx = h_i p(\bar{x}_i) + O(h_i^2),$$

$$\int_{E_i} q(x)[\varphi_i(x)]^2 dx = q(\bar{x}_i)h_i/4 + O(h_i^2),$$

$$\int_{E_i} q(x)\varphi_i(x)\varphi_j(x)dx = q(\bar{x}_i)h_i/4 + O(h_i^2),$$

$$\int_{E_i} f(x)\varphi_j(x)dx = h_i f(\bar{x}_i)/2.$$

一般地，所选择的Gauss求积公式的精度应等于或略大于有限元解所具有的精度即可，不必选用精度过高的求积公式。

若分划节点是从左向右（或从右向左）顺序编号的，则K是3对角矩阵，故可用追赶法解有限元方程(1.25)，这是理想的解法之一，因其计算工作量为$O(N)$(N是内节点个数或有限元方程中未知数的个数)。解得的是$u(x_i)$的近似值$u_h(x_i)$。最后还可以写出边值问题(1.13)在$[a, b]$上的有限元近似解表达式

$$u_h(x) = \sum_{i=1}^{N} u_i\varphi_i(x),$$

并可据此计算一切需要计算的函数值，如某单元节点或单元中某点处的位移值或导数值。

§5 二次元

通常情况下，线性元解$u_h(x)$和真解的误差（参见第五章§4）为$|u_h(x)-u(x)|_\infty = O(h^2)$，其中$|\cdot|_\infty$表示误差的最大模。如果$u(x)$的光滑程度较好，人们可以采用分段高次多项式函数（**高次元**）来近似解函数，从而提高逼近误差阶，例如u_h是分段二次多项式函数（**二次元**），则$|u_h(x)-u(x)|_\infty = O(h^3)$。

现讨论在一个单元$E=[x_{i-1}, x_i]$上构造二次Lagrange型

插值函数．为此，增设单元中点$x_{i-1/2}$，设$N_{i-1}(x)$、$N_{i-1/2}(x)$和$N_i(x)$分别为节点x_{i-1}、$x_{i-1/2}$和x_i处的单元节点基函数，它们都是二次多项式函数，则$u_{h,E}(x)$可由单元节点基函数的线性组合表示，且有

$$\begin{cases}u_{h,E}(x)=u_{i-1}N_{i-1}(x)+u_{i-1/2}N_{i-1/2}(x)+u_iN_i(x),\\ N_i(x_j)=\delta_{ij}, \quad N_i(x)\in P_2(x).\end{cases}$$

利用映射$\xi=h_{i-1}^{-1}(x-x_{i-1})$，将单元$[x_{i-1}, x_i]$映射为$[0, 1]$标准单元．这样，在$[x_{i-1}, x_i]$上讨论构造$N_{i-j/2}(x)(j=0, 1, 2)$的问题就转化为在$[0, 1]$区间上讨论构造基函数$N_{j/2}(\xi)$ $(j=0, 1, 2)$的问题了．

因$N_0(\xi)$是一个二次式且$N_0\left(\frac{1}{2}\right)=N_0(1)=0$，故$N_0(\xi)$必含有$\left(\xi-\frac{1}{2}\right)$和$(\xi-1)$因子，且有

$$N_0(\xi)=c\left(\xi-\frac{1}{2}\right)\left(\xi-1\right),$$

再利用条件$N_0(0)=1$，立即可得

$$N_0(\xi)=(2\xi-1)(\xi-1).$$

同理可得

$$N_{1/2}(\xi)=4\xi(1-\xi),$$

$$N_1(\xi)=\xi(2\xi-1).$$

这三个标准单元节点基函数图形参见图1.4．

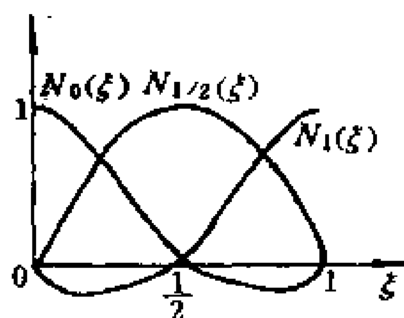


图 1.4

利用$\xi=h_{i-1}^{-1}(x-x_{i-1})$，$i=1, 2, \cdots, n$的逆变换，即可将标准区间$[0, 1]$上的节点基函数变换为$[x_{i-1}, x_i](i=1, 2, \cdots, n)$上的单元节点基函数，从而容易看出任意节点$x_i$和$x_{i-\frac{1}{2}}$处的整体节点基函数$\varphi_i(x)$和$\varphi_{i-\frac{1}{2}}(x)$的形状和特点，见图1.5．



图 1.5

这样，$u_h(x)=\sum_{i=0}^{2n} c_{\frac{i}{2}}\varphi_{\frac{i}{2}}$在各单元内部是二次多项式函数，在相邻两个单元的交接点处仅是连续的，即$u_h(x)$是分段二次可微且整体连续的函数。显然，$u_h(x)\in H^1$且能适合有限元方法的需要。

仿§4的讨论，即可将二次元用于有限元计算中，具体计算公式从略，若有必要，读者可以自行推导。

特别地，与(1.33)式相应的有限元方程组中那些和内部单元节点及单元中点有关的元素为

$$
\begin{cases}
a(\varphi_{i-1},\ \varphi_i)=\dfrac{1}{3}h_{i-1},\\
a(\varphi_{i-\frac{1}{2}},\ \varphi_i)=-\dfrac{8}{3}h_{i-1},\\
a(\varphi_i,\ \varphi_i)=\dfrac{7}{3}h_{i-1}+\dfrac{7}{3}h_i,\\
a(\varphi_{i+\frac{1}{2}},\ \varphi_i)=-\dfrac{8}{3}h_i,\\
a(\varphi_{i+1},\ \varphi_i)=\dfrac{1}{3}h_i,\\
a(\varphi_i,\ \varphi_{i+\frac{1}{2}})=-\dfrac{8}{3}h_i,\\
a(\varphi_{i+\frac{1}{2}},\ \varphi_{i+\frac{1}{2}})=\dfrac{16}{3}h_i,\\
a(\varphi_{i+1},\ \varphi_{i+\frac{1}{2}})=-\dfrac{8}{3}h_i,
\end{cases}
\qquad (1.35)
$$

$$F(\varphi_i)=\frac{1}{6}(h_{i-1}f_{i-\frac{1}{2}}+h_if_{i+\frac{1}{2}}),$$

$$F(\varphi_{i+\frac{1}{2}})=\frac{2}{3}h_if_{i+\frac{1}{2}}.$$

§6 关于提高有限元解精度的讨论

前面几节叙述了有限元方法的全过程及其计算公式，用有限元方法可求得位移函数$u_h(x)$在所有节点处的近似值。若采用线性元，则在各个单元上可计算位移

$$u_{h,E}(x)=u_{i-1}\frac{x_i-x}{h_{i-1}}+u_i\frac{x-x_{i-1}}{h_{i-1}},\qquad x\in E_{i-1},$$

及导数值（为常数）

$$u'_{h,E}(x)=(u_i-u_{i-1})/h_{i-1},\qquad x\in E_{i-1},$$

且理论上已证明

$$|u_{h,E}(x)-u_E(x)|_\infty=O(h^2),\qquad x\in E_i(i=0,\cdots,n-1),$$

$$|u'_{h,E}(x)-u'_E(x)|_\infty=O(h),\qquad x\in E_i(i=0,\cdots,n-1).$$

这些估计都是就单元整体而言的，是否在单元内部某些点上，位移近似值和导数近似值会出现较好的精度呢？这是**有限元超收敛性**研究中的问题，这里介绍部分实用结果。

对于线性元，位移$u_h(x)$在整个求解区间$[a,b]$上有二阶精度，即$|u(x)-u_h(x)|_\infty=O(h^2)$，其中节点值$u_1$，$u_2$，…，$u_n$的精度稍好些，但一般没有更高的阶；$u'_h(x)$在整个区间上一般只有一阶精度，即$|u'(x)-u'_h(x)|_\infty=O(h)$，但在每个单元的中点$\bar{x}_i=(x_{i-1}+x_i)/2$处，$u'_h(\bar{x}_i)$有二阶精度，即$|u'(\bar{x}_i)-u'_h(\bar{x}_i)|_\infty=O(h^2)$；利用相邻两单元中点的最佳导数值进行线性内插和外推，可得到其它点处的最佳导数值，例如在单元E_i和E_{i+1}的公共节点x_i处，利用内插得到

$$\bar{u}'_h(x_i)=\frac{h_{i+1}u'_h(\bar{x}_i)+h_iu'_h(\bar{x}_{i+1})}{h_i+h_{i+1}},$$

从而有

$$|u'(x_i)-\bar{u}'_h(x_i)|_\infty=O(h^2).$$

对于二次元，位移$u_h(x)$在整个求解区间$[a, b]$上一般有3阶精度，即$|u(x)-u_h(x)|_\infty=O(h^3)$；在单元端节点和单元中点处，二次元位移值的精度稍好些。二次元导数值也有超收敛性，在$\tilde{x}=\bar{x}_i\pm0.2886751346h_i$处，二次元导数值具有3阶精度，即$|u'(\tilde{x})-u'_h(\tilde{x})|_\infty=O(h^3)$；在单元内的其它点处，导数值只有2阶精度。当然也可以利用位移佳值或导数佳值作内插或外推，以获得其它点处的相应佳值。

利用有限元超收敛特性可有效地提高导数近似值的精度，若要有效地提高有限元位移值的精度，可采用**外推技术**。先对定解区间$[a, b]$作某种分划，例如等分n份，单元长度为h，并计算出有限元解$u_h(x)$。再将原分划细分，单元长度为$\frac{h}{2}$，也可计算出相应的有限元解$u_{h/2}(x)$。对线性有限元解，在节点x_i处利用外推公式

$$\tilde{u}_{h/2}(x_i)=\frac{1}{3}[4u_{h/2}(x_i)-u_h(x_i)],$$

则有

$$|\tilde{u}_{h/2}(x_i)-u(x_i)|_\infty=O((h/2)^3)\quad(i=1, 2, \cdots, n-1);$$

在单元中点$\bar{x}_i$处利用内插公式

$$\tilde{u}_{h/2}(\bar{x}_i)=\frac{1}{2}[\tilde{u}_{h/2}(x_i)+\tilde{u}_{h/2}(x_{i+1})],$$

则有

$$|\tilde{u}_{h/2}(\bar{x}_i)-u(\bar{x}_i)|_\infty=O((h/2)^3)\quad(i=1, 2, \cdots, n-1).$$

对于二次有限元解，在节点x_i处利用外推公式

$$\tilde{u}_{h/2}(x_i)=\frac{1}{7}[8u_{h/2}(x_i)-u_h(x_i)],$$

则有

$$|\tilde{u}_{h/2}(x_i)-u(x_i)|_\infty=O((h/2)^4)\quad(i=1, 2, \cdots, n-1);$$

在单元中点$\bar{x}_i$处利用内插公式

$$\tilde{u}_{h/2}(\bar{x}_i)=\frac{1}{2}[\tilde{u}_{h/2}(x_i)+\tilde{u}_{h/2}(x_{i+1})],$$

则有

$$|\tilde{u}_{h/2}(\bar{x}_i)-u(\bar{x}_i)|_\infty=O((h/2)^4) \quad (i=1,\ 2,\ \cdots,\ n-1)。$$

总之，利用有限元近似解的超收敛性、内插和外推技巧提高有限元近似解的精度，可以数十倍地节省计算工作量，这是一个值得进一步加深研究且值得在实用中推广的课题，有关这方面的知识，读者还可参见第四章§1，或参阅文献[9]、[14]、[19]~[23]。

第二章　解二维椭圆边值问题的有限元方法

在第一章中，我们已就一维椭圆边值问题讨论了有限元方法的基本思想及其实现过程。即先建立与微分方程等价的变分问题或变分方程；再构造由分段连续函数构成的有限维函数空间并利用Ritz-Galerkin 方法，便得到有限元方程组；解这个代数方程组就得到节点处的有限元近似解。这一思想过程同样也适用于二维情形，且更能体现出有限元方法解二维椭圆边值问题的优越性；但由于区域及其分划的特点，也随之产生了某些新问题。例如划分质量好坏严重地影响着有限元解的精度；节点编号不同，严重地影响着有限元方程系数矩阵的结构及求解有限元方程的效率等等。我们应当注意这些新的问题。

§1　二维椭圆边值问题及其等价的变分形式

1.1　二维椭圆边值问题

设Ω是$x-y$平面中的有界域，其边界$\partial\Omega$是分段光滑的闭曲线；记$\bar{\Omega}=\Omega\cup\partial\Omega$，$\partial\Omega=\Gamma_1\cup\Gamma_2$。考虑椭圆边值问题

$$\begin{cases} Lu=-(pu_x)_x-(pu_y)_y+ru=f,\ (x,\ y)\in\Omega, \\ u(x,\ y)=\alpha(x,\ y),\ (x,\ y)\in\Gamma_1, \\ p(x,y)\dfrac{\partial u}{\partial n}+\sigma(x,y)u=\beta(x,y),\ (x,y)\in\Gamma_2, \end{cases} \tag{2.1}$$

其中，$p(x,y)\in C^1(\bar{\Omega})$且$\min p(x,y)>0$；$r(x,y)\in C^0(\bar{\Omega})$且$r(x,\ y)\geqslant 0$；$\sigma(x,\ y)\in C^0(\bar{\Omega})$且$\sigma(x,\ y)\geqslant 0$；$f(x,\ y)\in C^0(\bar{\Omega})$；$\partial u/\partial n$表示$u$沿边界曲线的外法向导数(见图 2.1)，

$$\frac{\partial u}{\partial n}=\frac{\partial u}{\partial x}\cos(n,\ x)+\frac{\partial u}{\partial y}\cos(n,\ y);$$

$\beta(x,y)$和 $\sigma(x,y)\in C^0(\Gamma_2)$, $\sigma(x,y)\geqslant 0$; $\alpha(x,y)\in C^0(\Gamma_1)$; 如果 $\partial\Omega=\Gamma_2$ (即边界只由 Γ_2 组成) 则假定 r 与 σ 不同时为零。

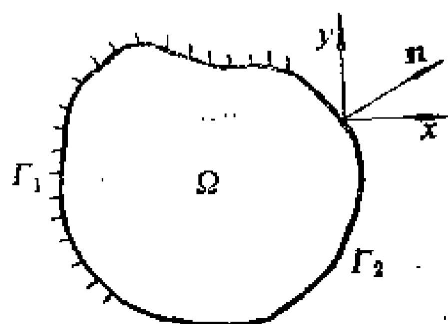


图2.1

上述条件都是为了保证边值问题(2.1)存在唯一解 $u\in C^2(\Omega)$。

如果系数 $p(x,y)\equiv 1$, $r(x,y)\equiv 0$, (2.1)式表示的偏微分方程就是 Poisson 方程

$$-\Delta u=-u_{xx}-u_{yy}=f(x,y),$$

它是最典型最重要的二阶椭圆型方程。许多恒定(驻定)的物理问题，例如薄膜的平衡问题，地下水的渗流问题，稳定温度场问题，静电场与磁场的势分布问题等等，这些物理现象虽有差别，但它们在数学上都可统一用边值问题(2.1)或 Poisson 方程边值问题表述。

1.2 扩充函数类

如同 §1 所述，边值问题(2.1)的解 $u\in C^2(\Omega)$ 称为经典解。为了在更广大的函数类中研究(2.1) 式相应的变分问题的解（弱解），我们引进二维的 Sobolev 空间(参见第五章§2)

$H^1(\Omega):=\{u(x,y)|\ u\in L_2(\Omega),\ u_x\in L_2(\Omega),\ u_y\in L_2(\Omega)\}$, 其模定义为

$$\|u\|_1^2:=\int_\Omega(u^2+u_x^2+u_y^2)dxdy.$$

根据定义，显然，连续可微的函数及在 Ω 上连续且分片可微的函数都属于 $H^1(\Omega)$。但是值得注意的是，对二维以上情形，$H^1(\Omega)$中的函数可能并不连续，这是与一个自变量情形不同的。例如在圆 $r=\sqrt{x^2+y^2}\leqslant\frac{1}{2}$ 上考虑函数 $u=(-\ln r)^\alpha$, $0<\alpha<$

1/2，利用极坐标，显然有

$$\int_{r<1/2} u^2 dxdy = 2\pi\int_0^{1/2} |\ln r|^{2\alpha} r dr < +\infty,$$

其一阶导数平方和的积分也有界，即

$$\int_{r<1/2}\left[\left(\frac{\partial u}{\partial x}\right)^2+\left(\frac{\partial u}{\partial y}\right)^2\right]dxdy = 2\pi\int_0^{1/2}\left(\frac{\partial u}{\partial r}\right)^2 rdr$$

$$= 2\pi\alpha^2\int_0^{1/2} |\ln r|^{2\alpha-2} r^{-1} dr$$

$$= \frac{2\pi\alpha^2}{1-2\alpha}(\ln 2)^{2\alpha-1} < +\infty,$$

因此 $u\in H^1(\Omega)$。但是，由

$$\lim_{\substack{x\to 0\\ y\to 0}} u(x, y) = \lim_{r\to 0} |-\ln r|^{\alpha} = +\infty$$

可见，$u=(-\ln r)^{\alpha}$ 在圆心处不连续。若要详细了解 Sobolev 空间中函数的“光滑程度”，有兴趣的读者可参见文献[5]和 [6]。

1.3 与二维椭圆边值问题等价的变分形式

与第一章讨论此问题的方法类似，可先形式地构造出与(2.1)式相应的变分问题和变分方程，然后再证明它们的等价性。为此，我们先在形式上观察

$$\int_{\Omega}(Lu-f)vdxdy,$$

式中，$v(x, y)$ 为试探函数，由(2.1)式中的边值条件可知，其基本特点是 $v(x, y)$ 在 Γ_1 边界上恒为零，在 Γ_2 上不被限制。利用 Green 公式

$$\int_{\Omega}-(pu_x)_x vdxdy = \int_{\Omega} pu_x v_x dxdy - \int_{\partial\Omega} pu_x \cos(n, x) vds,$$

$$\int_{\Omega}-(pu_y)_y vdxdy = \int_{\Omega} pu_y v_y dxdy - \int_{\partial\Omega} pu_y \cos(n, y) vds,$$

则有

$$\int_{\Omega}(Lu-f)vdxdy$$

$$
\begin{aligned}
&=\int_{\Omega}(pu_xv_x+pu_yv_y+ruv-fv)dxdy-\int_{\partial\Omega}p\frac{\partial u}{\partial n}vds\\
&=\int_{\Omega}(pu_xv_x+pu_yv_y+ruv-fv)dxdy-\int_{\Gamma_2}p\frac{\partial u}{\partial n}vds\\
&=\int_{\Omega}(pu_xv_x+pu_yv_y+ruv-fv)dxdy-\int_{\Gamma_2}(\beta-\sigma u)vds\\
&=\int_{\Omega}(p\nabla u\nabla v+ruv)dxdy+\int_{\Gamma_2}\sigma uvds-\Big[\int_{\Omega}fvdxdy\\
&\qquad\qquad+\int_{\Gamma_2}\beta vds\Big].
\end{aligned}
\tag{2.2}
$$

令(2.2)式中的二次泛函和线性泛函分别为

$$
a(u,\ v):=\int_{\Omega}(p\nabla u\nabla v+ruv)\,dxdy+\int_{\Gamma_2}\sigma uvds, \tag{2.3}
$$

$$
F(v):=\int_{\Omega}fvdxdy+\int_{\Gamma_2}\beta vds; \tag{2.4}
$$

并记 $H^1_E(\Omega):=\{u\in H^1(\Omega)\,|u|_{\Gamma_1}=\alpha\}$,

$H^1_{0E}(\Omega):=\{v\in H^1(\Omega)|\,v|_{\Gamma_1}=0\}$,

从而在形式上可建立与边值问题(2.1) 相应的变分问题和变分方程:

$$
J(u)=\frac{1}{2}a(u,\ u)-F(u)\Rightarrow\min,\qquad u\in H^1_E(\Omega), \tag{2.5}
$$

和

$$
a(u,\ v)=F(v),\ u\in H^1_E(\Omega),\ \forall v\in H^1_{0E}(\Omega). \tag{2.6}
$$

读者由 $a(u,\ v)$的定义(2.3)式不难看出，$a(u,\ v)$是一个对称的双线性泛函，(2.1)式解的唯一性表明了 $a(v,\ v)>0$，这个正定性证明见第五章§3，有兴趣的读者也可参阅文献［8］~［15］中的有关部分。

下面证明边值问题与相应构造出来的变分问题等价。

定理2.1 若$u^*\in C^2(\Omega)\cap H^1_E(\Omega)$是变分问题(2.5) 的解，则$u^*$也是边值问题(2.1)的解；反之，若$u^*$是边值问题(2.1) 的解，则$u^*$也是变分问题(2.5)的解。

证明 设$u^*\in C^2(\Omega)\cap H^1_E(\Omega)$是变分问题(2.5)的极小解，

那么令 $u=u^*+\lambda v$ （$\lambda\in R$，v 是试探函数），将其代入 $J(u)$ 的表达式，并利用$a(u, v)$的对称性、双线性和 $F(v)$的线性，就有

$$J(u)=J(u^*+\lambda v)=\frac{1}{2}a(u^*+\lambda v,\ u^*+\lambda v)-F(u^*+\lambda v)$$

$$=\frac{1}{2}a(u^*,u^*)-F(u^*)+\lambda[a(u^*,v)-F(v)]+\frac{\lambda^2}{2}a(v,v). \quad (2.7)$$

据变分法，则变分问题(2.5) 必满足取得极小解的必要条件和唯一性条件

$$J'_\lambda(u^*+\lambda v)\big|_{\lambda=0}=a(u^*,\ v)-F(v)=0, \quad (2.8)$$

和

$$J''_\lambda(u^*+\lambda v)\big|_{\lambda=0}=a(v,\ v)>0. \quad (2.9)$$

又因为$u^*\in C^2(\Omega)$，所以利用(2.8)式就有

$$\begin{aligned}a(u^*,\ v)-F(v)&=\int_\Omega(p\nabla u^*\nabla v+ru^*v)dxdy\\&\quad+\int_{\Gamma_2}\sigma u^*vds-\int_\Omega fvdxdy-\int_{\Gamma_2}\beta vds\\&=\int_\Omega[-(pu_x^*)_x-(pu_y^*)_y+ru^*-f]vdxdy\\&\quad+\int_{\partial\Omega}p\frac{\partial u^*}{\partial n}vds+\int_{\Gamma_2}(\sigma u^*-\beta)vds\\&=\int_\Omega[Lu^*-f]vdxdy+\int_{\Gamma_2}\left[p\frac{\partial u^*}{\partial n}+\sigma u^*-\beta\right]vds=0.\end{aligned}$$

再利用试探函数 v 的任意性（准确地说，利用变分预备定理，参见第五章§1)，就有

$$-\frac{\partial}{\partial x}\left(p\frac{\partial u^*}{\partial x}\right)-\frac{\partial}{\partial y}\left(p\frac{\partial u^*}{\partial y}\right)+ru^*=f,\qquad(x,\ y)\in\Omega,$$

$$p(x,\ y)\frac{\partial u^*}{\partial n}+\sigma(x,\ y)u^*=\beta(x,\ y),\qquad(x,\ y)\in\Gamma_2.$$

还因为$u^*\in C^2(\Omega)\cap H^1_E(\Omega)$，所以$u^*|_{\Gamma_1}=\alpha(x,\ y)$。故证得$u^*$也是边值问题(2.1)的解。

若u^*是边值问题 (2.1)的解且$u^*\in C^2(\Omega)\cap H^1_E(\Omega)$，则由(2.2)式知$u^*$满足变分方程(2.6)，又因$a(v,v)>0,\forall v\in H^1_{0E}(\Omega)$

(具体证明见第五章§3)，所以u^*满足变分问题(2.5)取得唯一极小解的必要条件(2.8)和唯一性条件(2.9)，故u^*也是变分问题(2.5)的极小解。 □

仿照定理2.1的部分证明，读者易证边值问题(2.1)和变分方程(2.6)的等价性，即有

定理2.2 若$u\in C^2(\Omega)\cap H_E^1(\Omega)$，则边值问题(2.1)和变分方程(2.6)等价。

应该注意，变分问题(或变分方程)的解$u\in H_E^1(\Omega)$，一般并非属于$C^2(\Omega)$，甚至可能是不连续的函数，因此它被称为相应边值问题的**广义解**或**弱解**。(2.6)式就是相应于边值问题(2.1)的**弱解方程**。只有变分形式的解足够光滑，即$u\in C^2(\Omega)\cap H_E^1(\Omega)$时，广义解和经典解才是一致的，定理2.1和定理2.2正反映了这种一致性。

变分问题取得极小解的必要条件是变分方程，唯一性条件是二次泛函$a(v, v)$在试探函数空间中正定。$a(v, v)$的正定性也是(2.1)式中微分算子的椭圆性的表现，也是边值问题(2.1)有唯一解的表现。正因为这个原因，我们总是先假定边值问题有唯一解，这样就可以毫不费力地建立起相应等价的变分形式(当然弱解也是唯一的，参见第五章§3)；然后再仿照第一章的讨论，利用有限元方法(有限元方程此时也是唯一可解的)解得边值问题的有限元数值解。如果不知道边值问题是否唯一可解，或从实际问题中直接建立起变分问题，那就需要在理论上证明微分算子的椭圆性或变分形式中二次泛函的正定性，以保证有限元方程唯一可解。

二维边值问题的有限元方法的基本思想也相仿于一维情形，即从变分问题或变分方程出发(变分形式在有限元方法中所起的作用如第一章§2所述)，再利用Ritz-Galerkin方法将变分方程离散，即获得一个在有限维空间中求极小解的有限元方程组，其关键在于构造出适合要求的有限维试探函数空间的基函数。

§2 三角线性元

用有限元方法解二维椭圆边值问题时，应用最广的是三角线性元。即将定解区域分划成若干个小三角形(称作**单元**)；在每个单元上构造插值型线性函数；用分片线性（每个单元上线性）但整体连续的函数近似代替变分问题或变分方程中的位移函数。

2.1 三角剖分与节点编号

将定解区域 $\bar{\Omega}$ 分划成若干个小**三角形单元** Δ_i，见图2.2*a*)、*b*)，每个单元的外接圆直径和内切圆直径相应记为 h_i 和 ρ_i，且记 $h=\max h_i$。

三角剖分时应注意以下几点：

1．为了方便构造插值型函数，要求每个单元的顶点一定是相邻单元的顶点，而不是相邻单元边上的点，见图2.2*a*)、*b*)、*c*)。

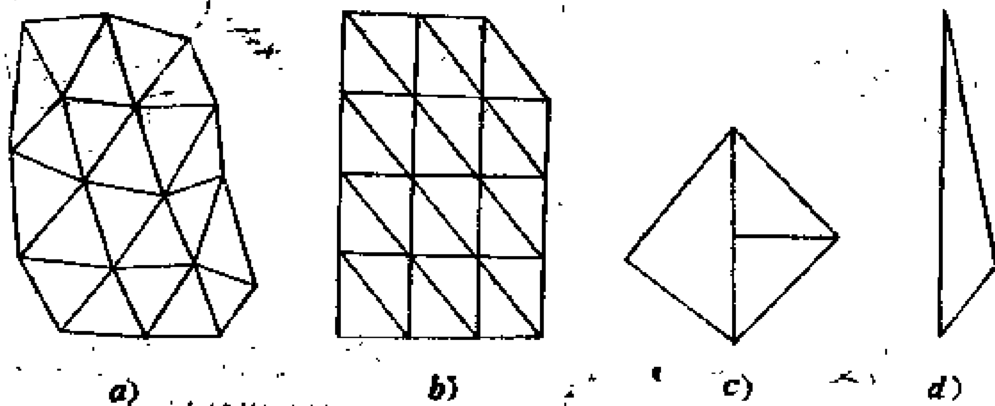


图2.2

2．为了保证有限元解有较好的精度，要求三角剖分是**"正规"**的，即要求每个单元中尽量避免出现大的钝角，见图2.2*d*)，且对一切单元来说 h_i/ρ_i 的上下界越接近越好。

对一般的定解区域很容易作出正规剖分，一切单元的集合 Ω_h 是区域 $\bar{\Omega}$ 的离散。连接所有落在边界 $\partial\Omega$ 上的节点的封闭曲线 $\partial\Omega_h$

可很好地逼近$\partial\Omega$。这就是说，三角形剖分具有很好的灵活性和适应性。

当然，三角剖分Ω_h可以是多种多样的，特别可根据实际问题的需要安排节点的位置。例如对扳手进行受力分析时，可根据光弹性力学的简单试验结果或根据经验，让扳手区域三角分划的节点安排更合理些，见图2.3。在位移或应力可能出现较大变化的区域，例如图2.3中等位线数值较高处，可适当加密节点；而

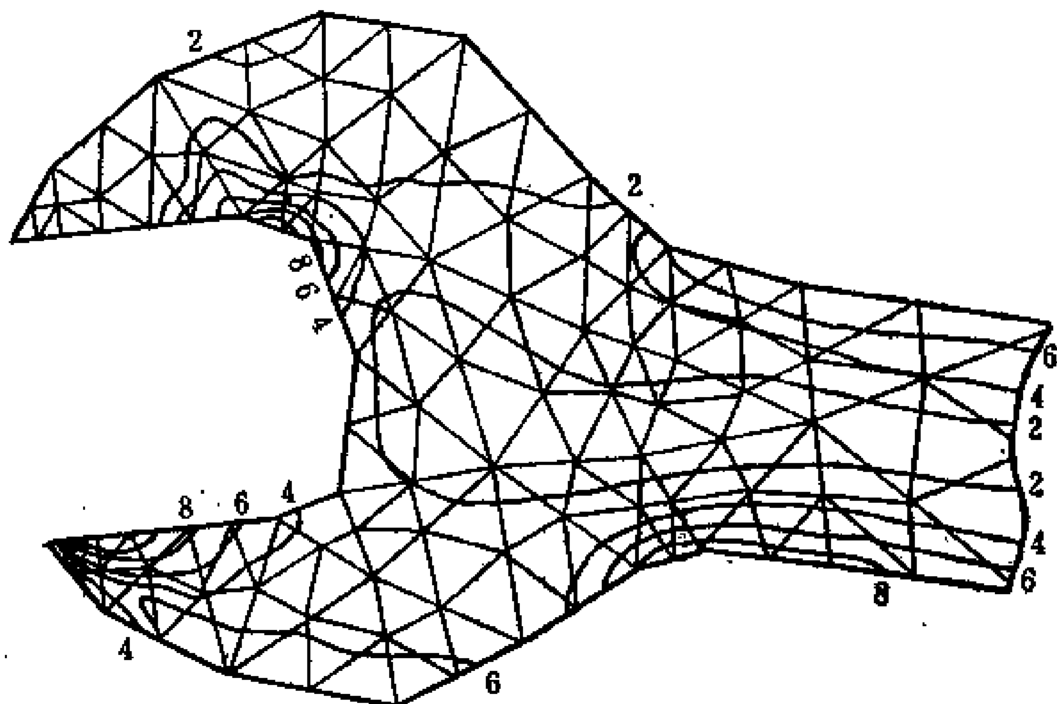


图2.3

位移或应力变化平缓处，可较少设置节点；相隔很远的单元之间，相应边长可相差数倍，但从稀到密应是逐渐过渡的，而不是急剧变化的。

再将Ω_h中一切节点编号，这种编号可以任意，但编号顺序将直接影响有限元方程组系数阵的结构；若用直接方法解有限元方程，则节点编号顺序还会影响存贮量和计算量；若用某种迭代法(如SOR方法)解有限元方程组，则存贮量和计算量可与节点编号顺序无关(参见第三篇)。

2.2 面积坐标及其性质

在三角形单元上构造插值型函数，并不简单类同于一维情形，它有自己的特点。为此，需要从新的角度进行研究。

考虑一个三角形单元，其面积计为S，其顶点按反时针顺序记为A_i、A_j和A_k，其中i、j和k为顶节点的编号。在此单元内部任取一点$P(x, y)$，连接P和三个顶点，此单元则被分成三个小三角形，它们的面积分别记为S_i、S_j和S_k(如图2.4)。记

$$L_i = S_i/S, \quad L_j = S_j/S, \quad L_k = S_k/S, \tag{2.10}$$

显然

$$\begin{cases} S_i + S_j + S_k = S, \quad L_i、L_j、L_k \geqslant 0, \\ L_i + L_j + L_k = 1, \\ L_m(A_n) = \delta_{mn} = \begin{cases} 1, & m = n, \\ 0, & m \neq n, \end{cases} \end{cases} \tag{2.11}$$

从而，单元内任一点P的位置可由一个三维数(L_i, L_j, L_k)所决定，P的坐标(x, y)和(L_i, L_j, L_k)是一一对应的。所以称(L_i, L_j, L_k)为**面积坐标**。

利用面积坐标的定义，可以很方便地确定单元中任一点的面积坐标，且可建立起一个面积坐标网。如图2.5所示，P_1点的面

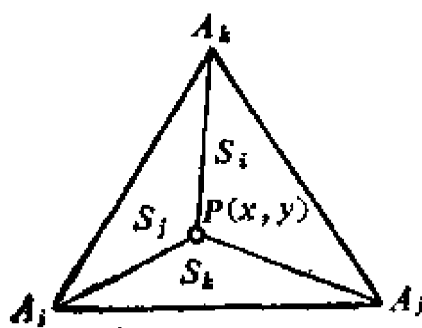


图2.4

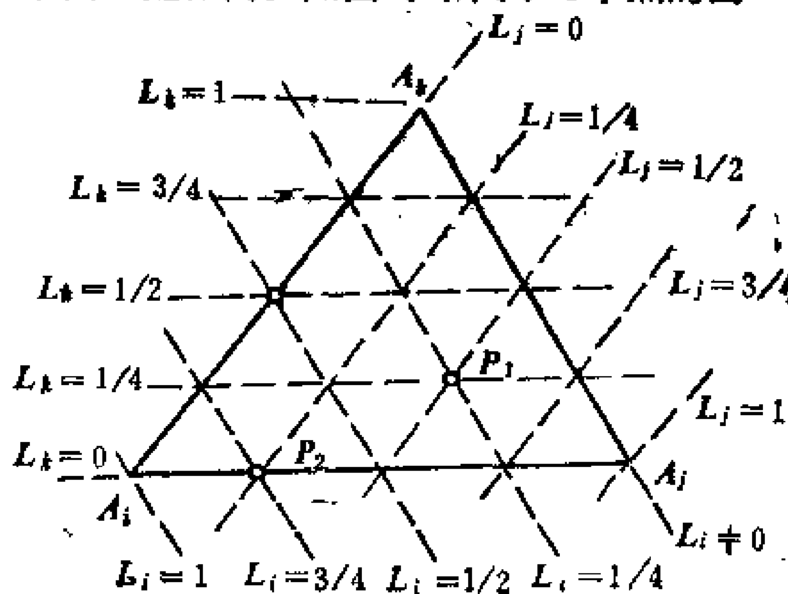


图2.5

积坐标是(1/4, 1/2, 1/4)，P_2 点的面积坐标是(3/4, 1/4,0)。以后将会看到，利用这个面积坐标网，构造插值型函数是非常方便的。

为了看清面积坐标和直角坐标之间的关系，我们有

$$2S=\begin{vmatrix}1 & x_i & y_i\\ 1 & x_j & y_j\\ 1 & x_k & y_k\end{vmatrix},\qquad 2S_i=\begin{vmatrix}1 & x & y\\ 1 & x_j & y_j\\ 1 & x_k & y_k\end{vmatrix},$$

$$2S_j=\begin{vmatrix}1 & x & y\\ 1 & x_k & y_k\\ 1 & x_i & y_i\end{vmatrix},\qquad 2S_k=\begin{vmatrix}1 & x & y\\ 1 & x_i & y_i\\ 1 & x_j & y_j\end{vmatrix};$$

$$\begin{cases}L_i=\dfrac{2S_i}{2S}=\dfrac{1}{2S}[(x_jy_k-y_jx_k)+(y_j-y_k)x+(x_k-x_j)y];\\ L_j=\dfrac{2S_j}{2S}=\dfrac{1}{2S}[(x_ky_i-y_kx_i)+(y_k-y_i)x+(x_i-x_k)y];\\ L_k=\dfrac{2S_k}{2S}=\dfrac{1}{2S}[(x_iy_j-y_ix_j)+(y_i-y_j)x+(x_j-x_i)y],\end{cases}\tag{2.12}$$

从而可得

$$\begin{cases}x=x_iL_i+x_jL_j+x_kL_k=x_i+(x_j-x_i)L_i+(x_k-x_i)L_j;\\ y=y_iL_i+y_jL_j+y_kL_k=y_i+(y_j-y_i)L_i+(y_k-y_i)L_j.\end{cases}\tag{2.13}$$

若记 $\xi=L_i$，$\eta=L_j$，则 (2.13) 式表明了一种变换，它把 x-y 坐标系中的任意三角形单元映射为 ξ-η 坐标系中的标准等腰直角三角形单元，如图 2.6 所示。这个变换的 Jacobi 行列式为

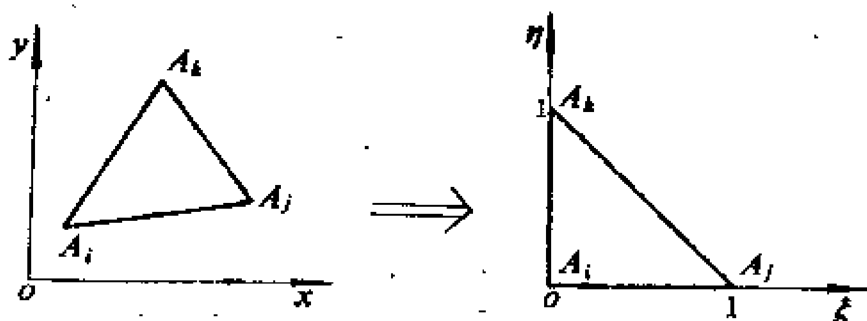


图2.6

$$J=\left|\frac{\partial(x,y)}{\partial(\xi,\eta)}\right|=\begin{vmatrix}\dfrac{\partial x}{\partial\xi} & \dfrac{\partial x}{\partial\eta}\\ \dfrac{\partial y}{\partial\xi} & \dfrac{\partial y}{\partial\eta}\end{vmatrix}=\begin{vmatrix}x_j-x_i & x_k-x_i\\ y_j-y_i & y_k-y_i\end{vmatrix}=2S. \qquad (2.14)$$

该变换是**仿射变换**，它除了能将三角单元仿射变换为**标准三角单元**外，还能将三角单元上的插值型函数(特别地，如基函数)变换为标准三角单元上的同类型函数。因此，利用面积坐标可使计算工作简单化、标准化。为此，我们列出几个以后要经常用到的公式。

由(2.12)式可得

$$\begin{cases}\dfrac{\partial L_i}{\partial x}=\dfrac{1}{2S}(y_j-y_k), & \dfrac{\partial L_i}{\partial y}=\dfrac{1}{2S}(x_k-x_j);\\ \dfrac{\partial L_j}{\partial x}=\dfrac{1}{2S}(y_k-y_i), & \dfrac{\partial L_j}{\partial y}=\dfrac{1}{2S}(x_i-x_k);\\ \dfrac{\partial L_k}{\partial x}=\dfrac{1}{2S}(y_i-y_j), & \dfrac{\partial L_k}{\partial y}=\dfrac{1}{2S}(x_j-x_i).\end{cases} \qquad (2.15)$$

由(2.13)和(2.14)两式可得

$$dxdy=Jd\xi d\eta, \qquad (2.16)$$

式中，J 由(2.14)式所定义，它反映了仿射变换的放大倍数。由(2.13)式分别对 x 和 y 求导，且注意复合函数微分法则，即得

$$\begin{cases}\dfrac{\partial\xi}{\partial x}=\dfrac{\partial y}{\partial\eta}\Big/J, & \dfrac{\partial\eta}{\partial x}=-\dfrac{\partial y}{\partial\xi}\Big/J,\\ \dfrac{\partial\xi}{\partial y}=-\dfrac{\partial x}{\partial\eta}\Big/J, & \dfrac{\partial\eta}{\partial y}=\dfrac{\partial x}{\partial\xi}\Big/J.\end{cases} \qquad (2.17)$$

利用面积坐标还有一个好处：任一关于 x 和 y 的 Q 次多项式都是关于面积坐标 L_i、L_j、L_k 的 Q 次齐次式，反之亦然。事实上，只需考虑关于 x 和 y 的 Q 次多项式中的任意一个单项式 $a_{mn}x^my^n(0\leqslant m+n\leqslant Q)$ 即可，由(2.11)和(2.13)两式即知

$$\begin{aligned}a_{mn}x^my^n&=a_{mn}x^my^n(L_i+L_j+L_k)^{Q-m-n}\\&=a_{mn}(x_iL_i+x_jL_j+x_kL_k)^m(y_iL_i+y_jL_j+y_kL_k)^n\\&\quad\cdot(L_i+L_j+L_k)^{Q-m-n}.\end{aligned}$$

这三个括号展开后，分别是关于 L_i、L_j、L_k 的齐 m 次、齐 n 次和齐 $Q-m-n$ 次多项式，故有

$$a_{mn}x^m y^n = b_{mn}L_i^p L_j^q L_k^r, \qquad p+q+r\equiv Q.$$

这就是说，任意定义在三角单元上的多项式函数（特别地如插值函数的基函数），在采用面积坐标后，形式简单且更便于设计和应用。

用面积坐标表示的齐次多项式在标准单元 $\bar{e}$ 上的积分也是极易计算的，可以利用下面的 Euler 积分公式：

$$\iint_{\bar{e}} L_i^p L_j^q L_k^r dL_i dL_j = \iint_{\bar{e}} L_i^p L_j^q L_k^r dL_i dL_k$$

$$=\iint_{\bar{e}} L_i^p L_j^q L_k^r dL_j dL_k = \frac{p!q!r!}{(p+q+r+2)!}. \qquad (2.18)$$

2.3 三角线性元

任一个三角单元上，可唯一确定一个线性插值函数

$$u_h(x, y) = \sum_{i=1}^{3} N_i(x, y)u_i,$$

其中 u_i 为三角形单元顶点 A_i 处给定的函数值；$N_i(x, y)$ 为 A_i 处相应的单元节点基函数，它们都是一次式。

利用面积坐标，任一个三角形单元被映射为一个标准单元，插值基函数则被映射为一个关于面积坐标的齐一次多项式，即

$$\begin{cases} N_i(L_1, L_2, L_3) = C_{i1}L_1 + C_{i2}L_2 + C_{i3}L_3; \\ N_i(A_j) = \delta_{ij}, \quad i、j=1, 2, 3, \end{cases}$$

其中 C_{ij}(i、$j=1, 2, 3$)为待定常数。由此式立即获得单元节点基函数

$$N_1 = L_1, \quad N_2 = L_2, \quad N_3 = L_3,$$

即面积坐标正好是某三角单元上线性插值函数的基函数。再回到三角形$A_1A_2A_3$上观察，即知其上的插值型函数可表示为

$$u_h(x, y) = \sum_{i=1}^{3} L_i u_i. \qquad (2.19)$$

在某个三角单元上线性插值基函数的图形如图 2.7a)、b)和

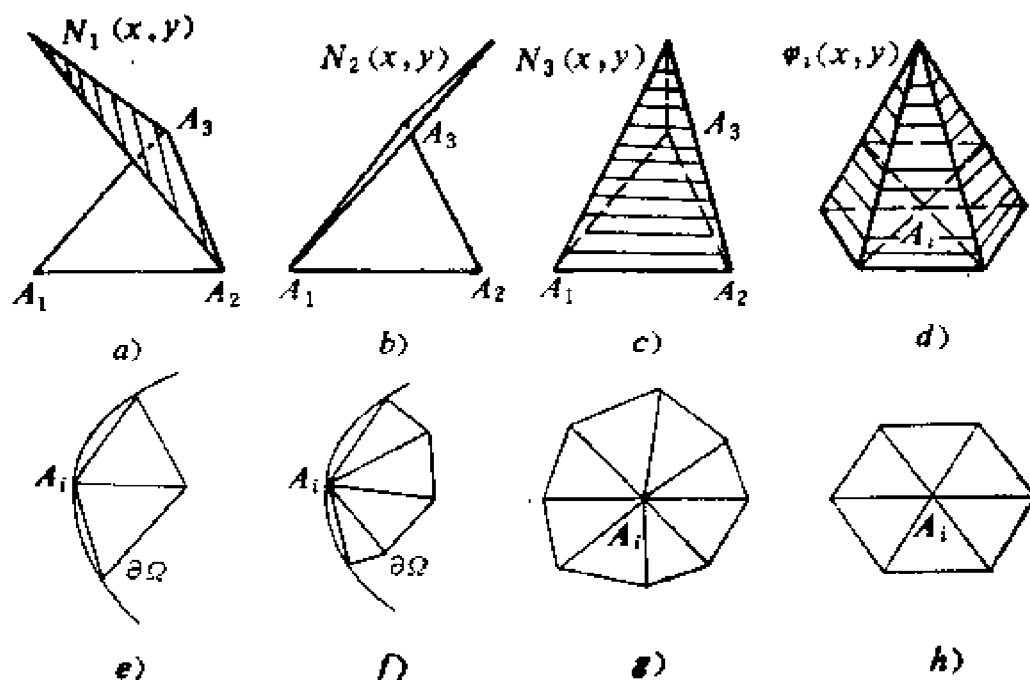


图2.7

c)所示。图2.7d)则表示在整个区域分划中编号为 i 的某节点 A_i 处的**整体节点基函数** $\varphi_i(x, y)$，φ_i 在 A_i 处为1，在以 A_i 为顶点的所有三角形上(这些三角形称为节点A_i 的**支集**，记为 supp A_i)φ_i 为线性函数，在 A_i 的支集外 φ_i 都等于零。图 2.7e)和 f)表示了 $A_i \in \partial\Omega$ 的支集，图2.7g)和 h)表示了$A_i \in \Omega$ 的支集。总之，整体节点基函数的特性可用下式表示：

$$\begin{cases} \varphi_i(A_j) = \delta_{ij} = \begin{cases} 1, & i=j, \\ 0, & i \neq j \text{ 或 } A_j \notin \text{supp } A_i, \end{cases} \\ \varphi_i(x, y) \in P_1(x, y), \quad (x, y) \in \text{supp } A_i. \end{cases} \tag{2.20}$$

Ω_h 中的每个节点都对应着一个整体节点基函数 $\varphi_i(x, y)$ $(i=1, 2, \cdots, N, N$ 为 Ω_h 中节点总数)，由这些整体节点基函数很容易构造出分片线性但整体连续的容许函数类 V_h：

$$V_h = \text{span}\{\varphi_1, \varphi_2, \cdots, \varphi_N\} \subset H_E^1(\Omega).$$

2.4 有限元方程组及其计算公式

利用(2.20)式定义的整体节点基函数，很容易构造出那些V_h

中满足各种边值条件的分片线性整体连续的函数。例如，设Γ_1^h是第一边值边界，那么作为容许函数$u(x, y)$的近似$u_h(x, y)$，可表示为

$$
\begin{cases}
u_h(x, y) = \sum_{i=1}^{N} \varphi_i(x, y) u_i; \\
u_h(A_i)\big|_{A_i \in \Gamma_1^h} = \varphi_i(A_i) \cdot u(A_i) = \alpha(x_i, y_i);
\end{cases}
\tag{2.21}
$$

作为试探函数$v(x, y)$的近似$v_h(x, y)$，可表示为

$$
\begin{cases}
v_h(x, y) = \sum_{j=1}^{N} \varphi_j(x, y) v_j; \\
v_h(A_j)\big|_{A_j \in \Gamma_1^h} = 0.
\end{cases}
\tag{2.22}
$$

利用 Ritz-Galerkin 方法，将(2.21)和(2.22) 式代人变分问题(2.5)或变分方程(2.6)，得到

$$
a(u_h, v_h) = F(v_h), \qquad u_h\big|_{\Gamma_1^h} = \alpha(x, y), \qquad v_h\big|_{\Gamma_1^h} = 0, \tag{2.23}
$$

或

$$
\sum_{i=1}^{N} a(\varphi_i, \varphi_j) u_i = F(\varphi_j), \qquad j = 1, 2, \cdots, N, \tag{2.24}
$$

其中

$$
\begin{cases}
a(\varphi_i, \varphi_j) = \int_{\Omega_h} (p(x, y) \nabla\varphi_i \nabla\varphi_j \\
\qquad\qquad + r(x, y)\varphi_i\varphi_j) dxdy + \int_{\Gamma_2^h} \sigma\varphi_i\varphi_j ds, \\
F(\varphi_j) = \int_{\Omega_h} f\varphi_j dxdy + \int_{\Gamma_2^h} \beta\varphi_j ds.
\end{cases}
\tag{2.25}
$$

且方程组(2.24)中，$a(\varphi_i, \varphi_j)$和$F(\varphi_j)$的数值要随不同节点的位置而改变，它还要反映不同边值的影响，其具体计算公式见本段内容的后半部分。

有限元方程组(2.24)是一个形如

$$
KU = F \tag{2.26}
$$

的线性代数方程组，其中

$$K=\begin{pmatrix} a(\varphi_1,\ \varphi_1) & \cdots & a(\varphi_N,\ \varphi_1) \\ \cdots & \cdots & \cdots \\ a(\varphi_1,\ \varphi_N) & \cdots & a(\varphi_N,\ \varphi_N) \end{pmatrix},\quad U=\begin{pmatrix} u_1 \\ \vdots \\ u_N \end{pmatrix},\quad F=\begin{pmatrix} F(\varphi_1) \\ \vdots \\ F(\varphi_N) \end{pmatrix},$$

K又称为**总刚度矩阵**，U称为**位移向量**，F为**总载荷向量**。

由$a(\cdot,\ \cdot)$的正定对称性，知有限元方程组系数矩阵K是正定对称矩阵，所以有限元方程组(2.24)或(2.26)式可解，解出的$(u_1,\ \cdots,\ u_N)$为位移函数u在节点处的近似值。$u(x,\ y)$的有限元近似解u_h可表示为

$$u_h(x,\ y)=\sum_{i=1}^{N}u_i\varphi_i(x,\ y).$$

当只考虑一个单元情形时，例如一个如图2.4所示的单元e，此单元上对应着三个单元节点基函数φ_i、φ_j和φ_k，就在这个单元上考虑有限元方程

$$\begin{pmatrix} a_{ii}^e & a_{ij}^e & a_{ik}^e \\ a_{ji}^e & a_{jj}^e & a_{jk}^e \\ a_{ki}^e & a_{kj}^e & a_{kk}^e \end{pmatrix}\begin{pmatrix} u_i^e \\ u_j^e \\ u_k^e \end{pmatrix}=\begin{pmatrix} F_i^e \\ F_j^e \\ F_k^e \end{pmatrix},$$

其中，$a_{ij}^e=a_e(\varphi_i,\ \varphi_j)$，$F_i^e=(f,\ \varphi_i)_e$，$a_e(\cdot,\ \cdot)$和$(\cdot,\cdot)_e$分别表示整体积分$a(\cdot,\ \cdot)$和$(\cdot,\ \cdot)$在单元$e$上的限制。这个单元$e$上的有限元方程可简记为　$K_eU_e=F_e$，

并称K_e、U_e和F_e分别为**单元刚度矩阵**(或**子刚阵**)、**单元位移向量**和**单元载荷向量**。

设Ω_h共有NE个单元，N个节点，则可以将K_e和F_e分别扩充成$N\times N$方阵和N维向量，而单元节点的编号序数i、j、k可按实际大小放在矩阵及向量的相应行列上。例如，当$i<j<k$时，

$$K_e=\begin{pmatrix} & \vdots & \vdots & & \\ \cdots a_{ii}\cdots & a_{ij}\cdots & a_{ik}\cdots \\ & \vdots & \vdots & & \\ \cdots a_{ji}\cdots & a_{jj}\cdots & a_{jk}\cdots \\ & \vdots & \vdots & & \\ \cdots a_{ki}\cdots & a_{kj}\cdots & a_{kk}\cdots \\ & \vdots & \vdots & & \end{pmatrix},$$

$$F_e = (\cdots F_i^e \cdots F_j^e \cdots F_k^e \cdots)^T,$$

就是单元刚度矩阵和单元载荷向量扩充后的结构，其中虚点表示零元素。最后把所有的单元上的有限元方程叠加：

$$\sum_{n=1}^{NE} K_{en} U_{en} = \sum_{n=1}^{NE} F_{en}$$

得到有限元方程(2.24)或(2.26)式，其中

$$K = \sum_{n=1}^{NE} K_{en}, \quad F = \sum_{n=1}^{NE} F_{en} .$$

上式表明总刚度矩阵和总载荷向量分别是单元刚度矩阵和单元载荷向量叠加的结果。这就是说，人们可以就每个单元先计算出单元刚度矩阵和单元载荷向量的数值，再经叠加，边值条件的处理，从而具体实现了总刚度矩阵各元素和总载荷向量各元素的计算问题。这种办法的结构清晰，计算过程很规则，目前介绍有限元方法的书籍中一般都是采用这种叠加的办法来形成有限元方程，就是某些国际上有名的有限元方法的通用程序也都是采用这种办法形成有限元方程。但作者认为，采用这种办法的结果，总刚度矩阵可采用变带宽存贮而不能采用非零元素压缩存贮（参见第三章），浪费了不少存贮单元，不利于微型机和小型机解决规模大的有限元计算问题。

本书将介绍另一种形成有限元方程的办法，这种办法有利于非零元素压缩存贮，存贮使用可达最佳状态，还有利于整个计算过程全自动化，尽可能地减少手工操作量。

下面讨论有限元方程(2.24)或(2.26)中诸积分的计算公式。由于基函数的局部非零特征，利用(2.12)、(2.15)和(2.20)式，易知有限元方程组中的诸积分都是与节点所处位置密切相关的，读者应注意这一点。

对内节点 j，当 j 的所有邻点 i 也都是内节点时，(2.25)式中的边界积分项为零，所以有

$$I = \int_{\Omega_n} [p(x, y)(\varphi_j)_x(\varphi_i)_x + p(x, y)(\varphi_i)_y(\varphi_j)_y] dx dy$$

$$+\int_{\Omega_h} r(x,\ y)\varphi_i\varphi_j dxdy$$

$$=\sum_{\mu}\int_{\Delta_\mu}\frac{1}{4S_\mu^2}p(x,y)[(y_j-y_k)(y_k-y_i)+(x_k-x_j)(x_i-x_k)]$$

$$\cdot dxdy+\sum_{\mu}\int_{\Delta_\mu}\frac{1}{4S_\mu^2}r(x,y)L_i(x,y)L_j(x,y)dxdy, \quad (2.27)$$

$$\text{II}=\int_{\Omega_h} p(x,y)[(\varphi_j)_x^2+(\varphi_j)_y^2]dxdy+\int_{\Omega_h} r(x,y)\varphi_j^2 dxdy$$

$$=\sum_{\mu}\int_{\Delta_\mu}\frac{1}{4S_\mu^2}p(x,y)[(y_k-y_i)^2+(x_i-x_k)^2]dxdy$$

$$+\sum_{\mu}\int_{\Delta_\mu}\frac{1}{4S_\mu^2}r(x,y)L_j^2(x,y)dxdy, \quad (2.28)$$

$$\text{III}=\int_{\Omega_h} f(x,y)\varphi_j dxdy=\sum_{\mu}\int_{\Delta_\mu} f(x,y)L_j(x,y)dxdy, \quad (2.29)$$

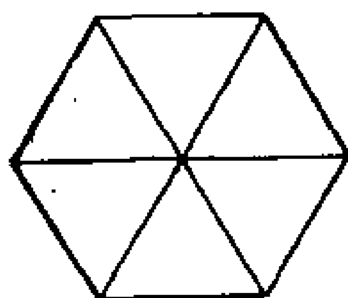
其中，$\sum\limits_{\mu}$表示在以$\overline{ji}$为边的所有三角单元Δ_μ上求和；当 $j=i$时，$\sum\limits_{\mu}$则表示在以 i 为顶点的一切三角单元 Δ_μ 上求和。

当 p、r 和 f 都是 x、y 的函数时，计算上面三个式子要用到数值积分，具体办法可参见第三章§2。当 p、r 都是常数时，不妨设为 1，积分式 I ～ III 的计算很简单，只要利用积分公式(2.18)，手工就可容易地算出：

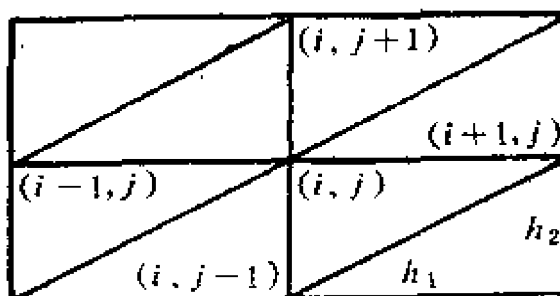
$$\begin{cases} \text{I}=\sum\limits_{\mu}\left[\dfrac{1}{4S_\mu}(\overrightarrow{jk}\cdot\overrightarrow{ki})+\dfrac{1}{12}S_\mu\right], \\ \text{II}=\sum\limits_{\mu}\left[\dfrac{1}{4S_\mu}(\overrightarrow{ki}\cdot\overrightarrow{ki})+\dfrac{1}{6}S_\mu\right], \\ \text{III}=\sum\limits_{\mu}\dfrac{1}{3}S_\mu f(x_j,\ y_j). \end{cases} \quad (2.30)$$

对于Poisson 方程和正三角形分划情形，参见图 2.8a)，利用(2.30)式不难算出

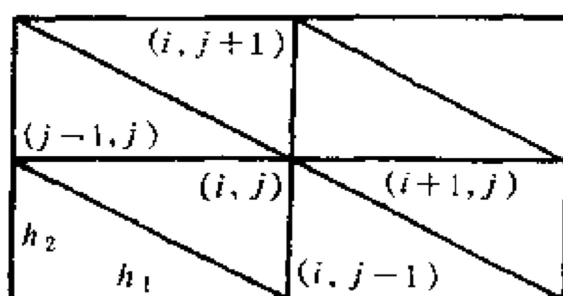
$$\begin{cases} a(\varphi_{j_l},\ \varphi_j)=-\sqrt{3}/3, \quad l=1,\ 2,\ \cdots,\ 6, \\ a(\varphi_j,\ \varphi_j)=2\sqrt{3}, \\ (f,\ \varphi_j)\text{可采用 4～6 阶精度的 Gauss积分计算。} \end{cases} \quad (2.31)$$



a)



b)



c)

图2.8

对于Poisson方程和直角三角形分划情形，参见图2.8b)、c)，h_1和h_2分别表示x和y方向的直角边长；设节点编号为坐标式编号，$(i,\ j)$的邻点中没有边界点；记$(i,\ j)$节点处的整体节点基函数为φ_{ij}；则利用(2.30)式容易得到与该节点对应的总刚阵元素和右端项元素为

$$\begin{cases} a(\varphi_{ij},\ \varphi_{i,\,j\pm1}) = -h_1/h_2, \ a(\varphi_{ij},\ \varphi_{i\pm1,\,j}) = -h_2/h_1; \\ a(\varphi_{ij},\ \varphi_{ij}) = 2(h_1/h_2 + h_2/h_1); \\ a(\varphi_{ij},\ \varphi_{i+1,\,j\pm1}) = a(\varphi_{ij},\ \varphi_{i-1,\,j\pm1}) = 0; \\ (f,\ \varphi_{ij}) = h_1 h_2 f_{ij}. \end{cases} \tag{2.32}$$

特别当 $h_1 = h_2 = h$ 时，有

$$\begin{cases} a(\varphi_{ij},\ \varphi_{i,\,j\pm1}) = a(\varphi_{ij},\ \varphi_{i\pm1,\,j}) = -1; \\ a(\varphi_{ij},\ \varphi_{ij}) = 4; \\ a(\varphi_{ij},\ \varphi_{i+1,\,j\pm1}) = a(\varphi_{ij},\ \varphi_{i-1,\,j\pm1}) = 0; \\ (f,\ \varphi_{ij}) = h^2 f_{ij}. \end{cases} \tag{2.33}$$

对内节点 j，当其邻点 i 有部分位于 Γ_1^h 上时，(2.25) 式中的边界积分项为零，但总刚阵元素要受到强制边界条件的影响。若设 j 的所有邻点中，编号为 $i1$ 到 im 的邻点是内节点，编号为 $im+1$ 到 $im+r$ 的邻点是第一边值节点，那么对应 j 节点处的有限元方程(2.24)可相应写为

$$\sum_{i=i1}^{im+r} a(\varphi_i,\ \varphi_j)u_i = F(\varphi_j).$$

因 $(x_i, y_i)\in\Gamma_1^h$ 有 $u_h(x_i, y_i)=u(x_i,\ y_i)=u_i=\alpha_i$，故上式可写成

$$\sum_{i=i1}^{im} a(\varphi_i,\ \varphi_j)u_i = F(\varphi_j) - \sum_{i=im+1}^{im+r} a(\varphi_i,\ \varphi_j)\alpha_i. \qquad (2.34)$$

(2.34)式表明了有限元方程的强制边值处理的特征。

对内节点 j，当其邻点 i 有部分位于 Γ_2^h 上时，(2.25)式中的边界积分项也为零，总刚阵元素的计算公式仍同于 (2.27)～(2.29)式。

若节点 j 是第一边值点，则在对应节点 j 处的有限元方程(2.24)或(2.26)中，应有

$$a(\varphi_j,\ \varphi_j)\equiv 1, \quad a(\varphi_i,\ \varphi_j)\equiv 0 \quad (i\neq j), \quad F(\varphi_j)=\alpha_j.$$

当节点 i 和 j 都位于 Γ_2^h 上时，如图2.9所示，(2.25)式中的区域积分项仍同于 Ⅰ、Ⅱ 和 Ⅲ，但(2.25)式的边界积分项不为零。此时 $\varphi_i(x,\ y)$ 和 $\varphi_j(x,\ y)$ 都被限制在 $\overline{ij}$ ($\subset\Gamma_2^h$)直线段上而变为一维基函数情形，因此可采用第一章§4中的有关计算公式，有

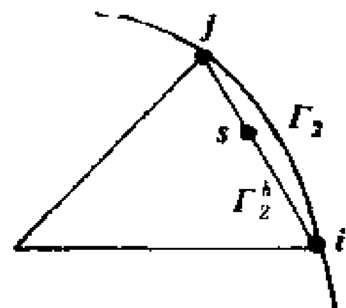


图2.9

$$\varphi_i(s)=|js|/|ij|,$$
$$\varphi_j(s)=|is|/|ij|.$$

从而 (2.25) 式的边界积分项计算公式为

$$\text{Ⅳ}=\int_{\Gamma_2^h}\sigma(s)\varphi_i\varphi_j\,ds=\sum_{i,j}\int_{\overline{ij}}\sigma(s)\frac{|js|\cdot|is|}{|ij|^2}\,ds, \qquad (2.35)$$

$$V=\int_{\Gamma_2^h}\beta(s)\varphi_j ds=\sum_{i_j}\int_{\overline{ij}}\beta(s)\frac{|is|}{|ij|}ds, \tag{2.36}$$

式中$\sum_{i_j}$表示在一切与 j 的邻点 i 的连线$\overline{ij}\subset\Gamma_2^h$的那些直线段上求和。

总之，对分划 Ω_h 的各个节点进行循环判断和计算，便可有效地处理各种边值并形成有限元方程组(2.24)。对此，读者还可参见第三章§4中的表4。

最后，根据问题的规模和计算机的容量、速度，选择适当的解法去求解有限元方程组；并编制程序上机计算。关于有限元程序设计及其过程中应该注意的问题，将在第三章和第四章介绍；关于有限元方程的高效率高精度解法，将在第三篇详细讲述。

第三章　有限元程序设计中的几个问题

本章主要讨论有限元程序设计的总刚阵(**大型稀疏矩阵**)存贮、网格自动生成及自动生成有限元方程组等方面的处理办法，还提供了计算框图。据此，读者可以编制利用有限元方法求解实际问题的通用程序。

§1　总刚阵结构及其压缩存贮方法

由于有限元方法采用了分片可微的整体节点基函数，所以这些基函数的局部非零特性就决定了总刚阵具有大型稀疏的特征。事实上，对于总刚阵元素 $a(\varphi_i,\varphi_j)$ 只是当节点 i 是节点 j 的邻点时，才可能 $a(\varphi_i,\varphi_j)\neq 0$，否则 $a(\varphi_i,\varphi_j)=0$。换句话说，总刚阵的某一行中，只有几个元素非零。特别地，当网格剖分细密、总刚阵阶数很大时，总刚阵的稀疏性表现得越特出。另一方面，总刚阵的结构(即非零元素的排列方式)依赖于分划节点的编号，例如图 3.1 和图3.5b)是同一区域的剖分但编号不同，图3.2a)和图3.2b) 分别是与之相应的矩阵结构。因此，二维情形的总刚阵结构要比一维情形复杂得多。

值得指出的是，当网格剖分特殊时，例如矩形剖分、直角三角形剖分或正三角形剖分等，总刚阵元素数值的规则性很强，参见(2.30)～(2.33)式，此时总刚阵元素虽然分布复杂但不必存贮。当剖分单元不规则时，必须

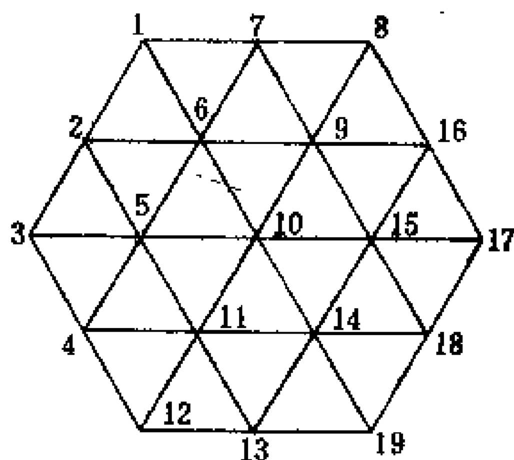
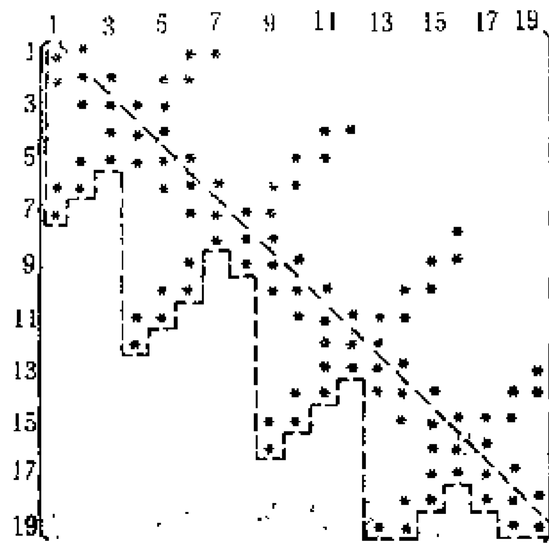


图3.1

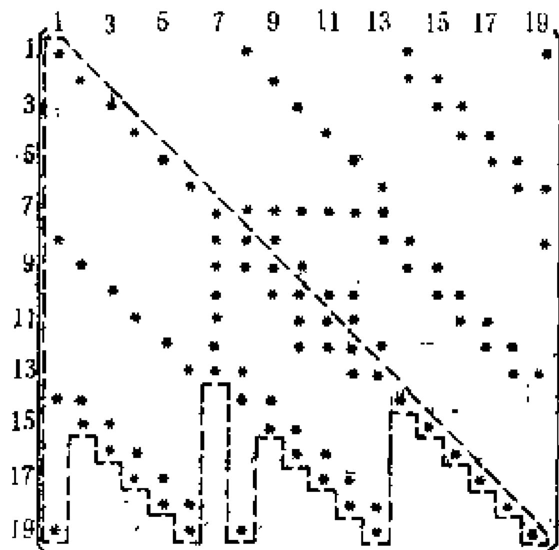
将总刚阵元素作有效地计算、存贮和查找，以便在求解有限元方程组时使用。

1.1 变带宽存贮方法

若采用直接解法（以消元法为基础，参见[18、19]）解有限元方程时，在消元过程中，例如从最后一行起按行消去矩阵的下三角部分，那么图3.2*a*)或图3.2*b*)中虚线框及其对称部分内的矩阵元素都要发生变化，而虚线框及其对称部分外的矩阵元零(全为零)不会参加消元过程。因此，采用直接解法解有限元方程时，考虑到对称性，必须将虚线框内的矩阵元素存贮起来（这称为**变带宽存贮方法**），尽管其中仍有不少零元素。



a)



b)

图3.2

此时，要存贮的总刚阵对角线以下的变带宽部分的元素可按列(也可按行)存放于一维数组中，为了查找和使用方便还需同时存贮对角元素的次序号。例如对6阶方阵K采用变带宽存贮，用于(按列)存放矩阵元素的一维数组是

$$E=(k_{11},k_{21},k_{22},k_{32},k_{33},k_{43},k_{53},k_{63},k_{44},k_{55},k_{65},k_{66}),$$

存放对角元素次序号的一维数组是

$$K=\begin{pmatrix} k_{11} & & & & & \text{对称} \\ k_{21} & k_{22} & & & & \\ & k_{32} & k_{33} & & & \\ & 0 & & k_{44} & & \\ & 0 & & \cdots & k_{55} & \\ 0 & k_{63} & & & k_{65} & k_{66} \end{pmatrix}$$

$$D=(1,3,5,9,10,12),$$

$D(i)$反映了对角元素 k_{ii} 在E中的位置。

若要寻找 k_{53}，可先找出 $D(3)=5$，则 k_{53} 必存贮在E中的第 $5+(5-3)$的位置上，即 $E(7)=k_{53}$。一般地，若 $i\geqslant j$，则有 $k_{i,j}=E(t)$，$t=D(j)+(i-j)$。

为了节省变带宽存贮的存贮量，就得尽量减少零元素的存贮量，因此必须设法使总刚阵中的非零元素尽量集中在对角线两侧。这就需要对节点编号进行优化处理。实践证明，下述节点编号的优化处理办法是较好的[24]。

第一步　首先人为地选取一个起始点，并给定编号“1”。要求起始点具有最小的“阶”(这里的“阶”是指一个节点与相邻节点连线的条数)，起始点一般在边界上。

第二步　将起始点的所有相邻节点编为第一“层”，并对第一“层”中的所有节点按节点“阶”数的增加依次进行累加编号。如此“层”中有几个节点具有相同的“阶”数，则编号顺序无特殊要求，可顺次编号。

第三步　与第一“层”中的所有节点相邻的一切未重新编号的节点编为第二“层”，对此“层”中的元素编号办法同于第二步。

第四步　如此继续下去，直到所有的节点都被重新编号为止。

1.2 非零元素压缩存贮方法

解有限元方程组，若采用迭代解法(例如 SOR 方法、Jacobi 迭代等)，或采用存取和计算过程类似于迭代法的算法(例如共轭斜量方法等)，或采用**多层网格方法**(该算法也是以松弛迭代为基础，它收敛速度快，计算工作量小且所需存贮单元少，是解有限元方程的高效率算法，具体参见第三篇)，在计算过程中总刚阵中的零元素不起作用，所以只要存贮那些非零元素就可以了，从而使存贮量达最佳状态，这就可能在一个较小内存的计算机中计算较大型的离散方程。特别值得强调的是，使用上述算法解有限元方程所需存贮单元的多少是与分划网格节点的编号顺序无关的，这样便可让机器自动进行网格剖分，也不需要进行节点编号的优化处理了。

这里将介绍两种非零元素的压缩存贮方法，读者可自行选择使用。

在非零元素压缩存贮方法中，非零元素可按行或列排成一个一维序列(叫做元素序列)；为指明每个非零元素在矩阵中的位置，还要一个一维序列存贮有关行或列的信息(称为地址序列)。这样，不仅只需存贮非零元素，最大可能地减少存贮量；而且可利用地址序列非常方便地在元素序列中查找所需非零元素。

设有一对称矩阵 A，其中写有 a_{ij} 的元素都是非零元素，其余的全为零元素。

$$A=\begin{pmatrix} a_{11} & & & & & \\ a_{21} & a_{22} & & \text{对称} & & \\ & & a_{33} & & & \\ a_{41} & & a_{43} & a_{44} & & \\ & a_{52} & & a_{54} & a_{55} & \\ & & & a_{64} & a_{65} & a_{66} \end{pmatrix}$$

方法一 利用A的对称性，把主对角线及其以下的非零元素

按行（也可按列）排成一个序列（即一维存贮）：

$$E=(a_{11},a_{21},a_{22},a_{33},a_{41},a_{43},a_{44},a_{52},a_{54},a_{55},a_{64},a_{65},a_{66}).$$

非零元素的查找由B和C决定：

$$B=(1,1,2,3,1,3,4,2,4,5,4,5,6),$$

$$C=(1,2,4,5,8,11),$$

其中，B中元素与E中元素一一对应，且表示E中元素位于A阵中的列数；C中元素分别代表A阵每行第一个非零元素在E中的位置；利用B和C就可确定任一非零元素a_{ij}在E中的位置。例如$a_{43}\neq0$，由于a_{43}必处于A的第4行第1个元素和第5行第1个元素之间，所以，首先由$C(4)=5$，$C(5)=8$知a_{43}一定位于E中第5个元素和第8个元素之间；由于E、B元素一一对应且B中元素表示E中元素在A阵中的列号，所以又可以在B中第5到第8个元素之间找到a_{43}的列号3，得$B(6)=3$；最后就可确定$E(6)=a_{43}$，即E中第6个元素是a_{43}。

方法二 元素序列E仍同上，地址序列由

$$B=(1,2,3,6,7,9,10,12,14,15,19,20,21)$$

给出，序列B与E的元素一一对应，B中每个元素的值表示E中对应元素在A阵下三角部分按行排列的位置，它们的对应关系是

$$d(i,j)=\frac{1}{2}(i-1)i+j, \tag{3.1}$$

其中i和j表示E中任一非零元素a_{ij}于A阵下三角部分的行号和列号（$i\geqslant j$）。因任一非零元素a_{ij}（$i\geqslant j$）在B中对应于一个数值$d(i,j)$，且B、E元素一一对应，所以由$d(i,j)$在B中的位置即可确定a_{ij}在E中的存放位置，例如$a_{43}\neq0$，先由$i=4$，$j=3$，即有$d(4,3)=9$；再找出数值9处于B中第6个元素位置，即$B(6)=9$；最后可确定E中第6个元素应是a_{43}，即$E(6)=a_{43}$。使用该方法时，每确定并计算出一个非零元素a_{ij}，

就顺次存放于 E 中，且利用公式(3.1)算得的数值 $d(i,j)$ 也顺次存放于 B 中；需要时，仍利用公式(3.1)在 E 中找出 a_{ij} 的数值。

比较这两种压缩存贮方法，显见方法一比方法二多用了N个存贮单元，但节省了查寻矩阵元素的时间；方法二则花费一定的计算时间来换取存贮空间的节省。

§2 数值积分

在计算总刚阵和右端项元素时，例如 (2.27)~(2.29) 等公式，常常要计算形如$\int_{\Delta}\Phi(x,y)dxdy$的积分值，对此通常用 Gauss 求积公式(参见[18]、[19])：

$$\int_{\Delta}\Phi(x,y)dxdy\approx S_{\Delta}\sum_{i}\omega_{i}\Phi(L_1^{(i)},L_2^{(i)}L_3^{(i)}),\qquad(3.2)$$

式中，S_{Δ} 表示某个单元△的面积，$(L_1^{(i)},L_2^{(i)},L_3^{(i)})$表明了△单元上的 Gauss **求积节点**的位置，$\Phi(L_1^{(i)},L_2^{(i)},L_3^{(i)})$表示$\Phi(x,y)$在求积节点处的函数值，$\omega_i$ 是求积系数。下面的表一中列出了几个常用精度的 Gauss 求积公式中所需要的求积系数和节点坐标。公式(3.2)中积分节点的面积坐标是对称分布的，若表 1 中出现 L_1、L_2、L_3的值有三个互异，则经过置换应有 6 个求积节点；若表 1 中出现的L_1、L_2、L_3的值有两个互异，则经置换后应有 3 个求积节点；点$\left(\frac{1}{3},\ \frac{1}{3},\ \frac{1}{3}\right)$是△单元的形心，它取作求积节点时只出现一次。

例如，与图 3.3④ 对应的 Gauss 积分式中有 4 个积分点，即 $P_1(0.3,0.3,0.3)$、$P_2(0.6,0.2,0.2)$、$P_3(0.2,0.6,0.2)$ $P_4(0.2,0.2,0.6)$，那么据求积公式(3.2)就有

$$\int_{\Delta}\Phi(x,y)dxdy=S_{\Delta}[0.5208333333(\Phi(P_2)+\Phi(P_3)+\Phi(P_4))-0.56250\Phi(P_1)]+O(h^4).$$

表3.1 三角形上数值积分公式(3.2)的积分系数和积分节点表

ω_i	L_1	L_2	L_3	误差	说明
1.00000 00000	0.33333 33333	0.33333 33333	0.33333 33333	$O(h^2)$	参见图3.3 ①
0.33333 33333	1.00000 00000	0.00000 00000	0.00000 00000	$O(h^2)$	参见图3.3 ②
0.33333 33333	0.55555 55555	0.55555 55555	0.00000 00000	$O(h^3)$	参见图3.3 ③
−0.56250 00000	0.33333 33333	0.33333 33333	0.33333 33333	$O(h^4)$	参见图3.3 ④
0.52083 33333	0.60000 00000	0.20000 00000	0.20000 00000		
0.16666 66667	0.65902 76224	0.23193 33686	0.10903 90091	$O(h^4)$	参见图3.3 ⑤
0.10995 17437	0.81684 75929	0.09157 62135	0.09157 62135	$O(h^5)$	参见图3.3 ⑥
0.22338 15897	0.10810 30182	0.44594 84909	0.44594 84909		
0.37500 00000	0.33333 33333	0.33333 33333	0.33333 33333	$O(h^5)$	参见图3.3 ⑦
0.10416 66667	0.73671 24990	0.23793 23665	0.02535 51346		
0.22503 00003	0.33333 33333	0.33333 33333	0.33333 33333	$O(h^6)$	参见图3.3 ⑧
0.12593 91805	0.79742 69854	0.10128 65073	0.10128 65073		
0.13239 41528	0.47014 20641	0.47014 20641	0.05971 58718		

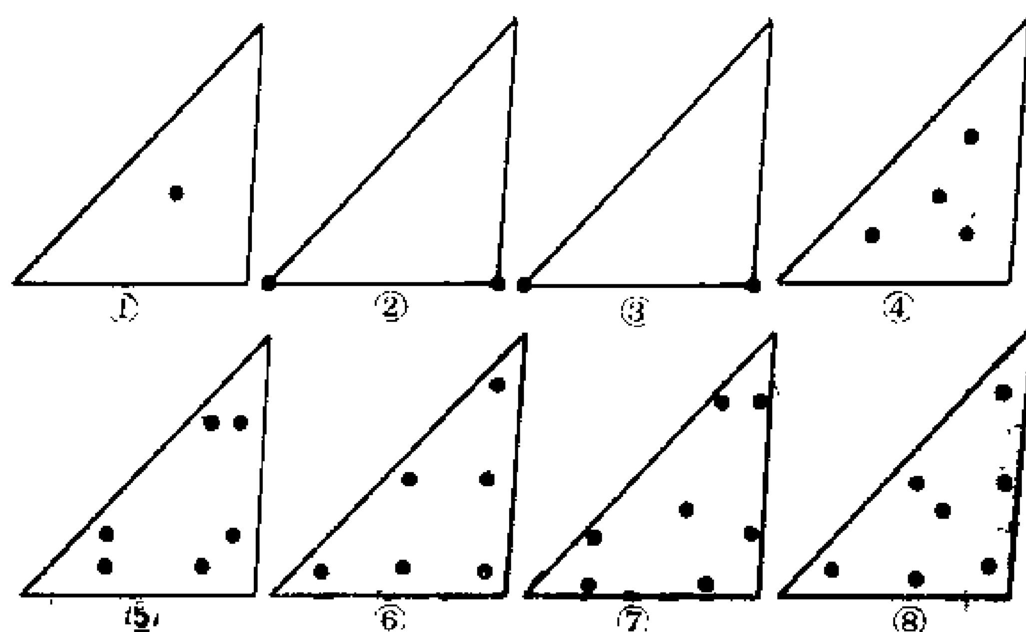


图3.3

在实际计算中，一般选取的高斯求积公式的精度应等于或略大于有限元解的精度。对于通常的正规三角分划，因线性有限元解的精度为 $O(h^2)$，所以可取精度为 $O(h^2)$的积分公式。例如一个三角单元的三个顶点记为P_1、P_2、P_3，形心记为 P，单元面积记为 S_Δ，那么可采用求积公式

$$\int_\Delta \Phi(x,y)dxdy = \Phi(P)S_\Delta + O(h^2),$$

或$\int_\Delta \Phi(x,y)dxdy = S_\Delta[\Phi(P_1)+\Phi(P_2)+\Phi(P_3)]/3+O(h^2)$。

将这种求积公式具体应用到（2.27）式中的积分项$\int_\Delta r(x,y)$ $L_i(x,y)L_j(x,y)dxdy$ 也是很容易的，此时形心和三顶点求积公式分别是

$$\int_\Delta r(x,y)L_iL_jdxdy = S_\Delta[r(P)L_i(P)L_j(P)]+O(h^2),$$

和

$$\int_{\Delta} r(x,y)L_iL_jdxdy=\frac{S_\Delta}{3}\sum_{l=1}^{6} r(P_l)L_i(P_l)+L_j(P_l)+O(h^2).$$

对于正三角形分划线性元，因有限元解的精度为 $O(h^4)$，所以要采用精度为 $O(h^4)$的求积公式。

对于有关线积分，一般可采用中点求积公式

$$\int_l \Phi(s)ds=|l|\Phi\left(\frac{P_1+P_2}{2}\right)+O(|l|^2),$$

或用梯形求积公式

$$\int_l \Phi(s)ds=|l|[\Phi(P_1)+\Phi(P_2)]/2+O(|l|^2),$$

其中 l 表示 $\overline{P_1P_2}$线段。

一般地，可编制具有不同精度的数值积分子程序，以便在形成有限元方程时选用。

§3 区域机器剖分

在形成有限元方程的计算过程中，需要提供哪些节点是内点还是某种类型的边界点；某一节点有哪些邻点；如何组成各个单元；每个单元顶点按反时针排列的顺序等信息。这些信息量是相当大的，特别当网格剖分细密时更是如此，若用手工操作输入这些信息，不仅艰难，而且还常常会出错，因此这些工作应该最大限度地让机器去完成。为此需输入一个节点信息表

节点号	邻点表	边值属性	阶数	x 坐标	y 坐标
I	LB(I,J)	LD(I)	LC(I)	X(I)	Y(I)

其中，阶数表示某节点所有相邻节点的个数，它可以反映该节点的支集中单元的个数；边值属性可反映该节点是内点、第一边值点还是第二、三边值点；邻点要求按反时针(或正时针)方向排列，不足最大阶数的位置用零补充，由此可确定该节点与哪些邻

点组成一个单元。总之，节点信息表提供了一切形成有限元方程所需要的信息。

手工编制较细网格层上的节点信息表往往也是一件很难很繁的工作，这一工作完全可以，也应该让机器来完成，这就是网格细分划和机器自动形成节点信息表的问题。不失一般性，设Ω是一个多边形区域，先给定一个初始的但是正规的三角分划，用手工编制并输入这个初始分划的节点信息表(其工作量是很小的)；然后，机器在执行过程中将每一个单元取边中点连结，形成一个较细一点儿的网格，要在初始网格节点编号的基础上累加节点编号，还要形成该网格层的节点信息表。如此继续，便可得到满足要求的网格层及其用于形成有限元方程的节点信息表。

对于曲线边界情形，机器自动剖分也可顺利实现，参见图3.4 *a*)、*b*)和*c*)。对于初始分划图 3.4 *a*)，取 $\overline{AB}$ 垂直平分线与$\widehat{AB}$的交点并将其与$\overline{AC}$和 $\overline{BC}$ 的中点连结，即可得较细分划图3.4 *b*)，再细分一次即可得图3.4*c*)。

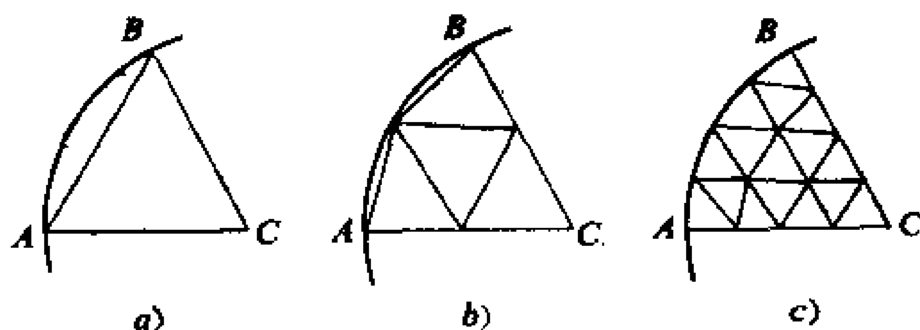


图3.4

下面举例说明机器自动剖分的过程，设有一个正六边形区域，其边界点都是第一边值点，将其初始分划成六个正三角形单元并给定节点编号，从而可手工编制一个初始的节点信息表（见图3.5 *a*)和表 2)，经机器细分一次和新添加点累加编号后可得

图3.5 b)和表3。如此下去可得到更细的分划和所需要用于计算的一切必要信息。

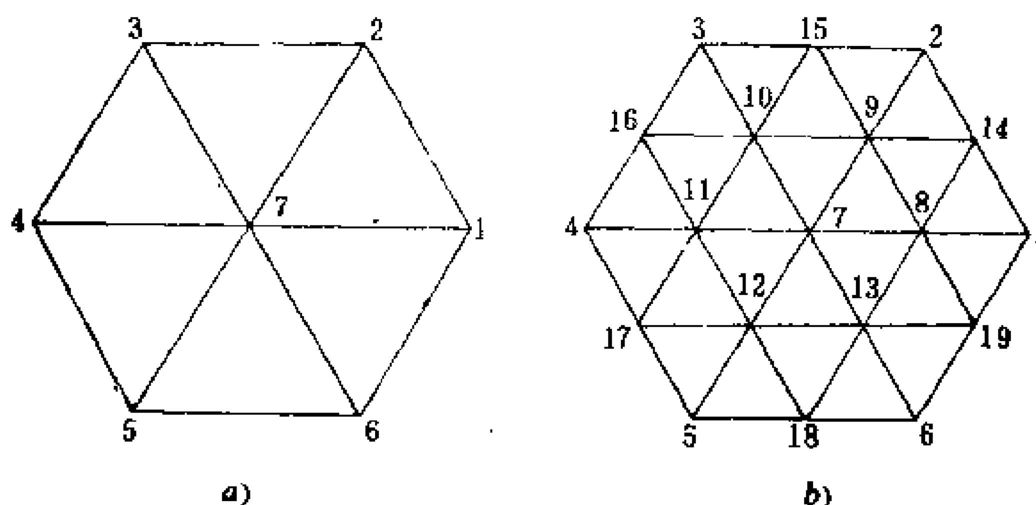


图3.5

为了对单连通区域、多连通区域，多边形区域、曲边形区域都采用机器自动剖分；为了适应各种边值问题（第一、第二或第

表3.2　对应图3.5a)分划的节点信息表

LB(7,6)	LD(7)	LC(7)	X(7)	Y(7)
2 7 6 0 0 0	1	3	1.00000	0.00000
3 7 1 0 0 0	1	3	0.50000	0.86603
4 7 2 0 0 0	1	3	－0.50000	0.86603
5 7 3 0 0 0	1	3	－1.00000	0.00000
6 7 4 0 0 0	1	3	－0.50000	－0.86603
1 7 5 0 0 0	1	3	0.50000	－0.86603
1 2 3 4 5 6	0	6	0.00000	0.00000

表3.3 对应图3.5b)分划的节点信息表

LB(19,6)						LD(19)	LC(19)	X(19)	Y(19)
14	8	19	0	0	0	1	3	1.00000	0.00000
15	9	14	0	0	0	1	3	0.50000	0.86603
16	10	15	0	0	0	1	3	−0.50000	0.86603
17	11	16	0	0	0	1	3	−1.00000	0.00000
18	12	17	0	0	0	1	3	−0.50000	−0.86603
19	13	18	0	0	0	1	3	−0.60000	+0.86603
8	9	10	11	12	13	0	6	0.00000	0.00000
1	14	9	7	13	19	0	6	0.50000	0.00000
2	15	10	7	8	14	0	6	0.25000	0.43301
3	16	11	7	9	15	0	6	−0.25000	0.43301
4	17	12	7	10	16	0	6	−0.50000	0.00000
5	18	13	7	11	17	0	6	−0.25000	−0.43301
6	19	8	7	12	18	0	6	+0.25000	−0.43301
2	9	8	1	0	0	1	4	0.75000	0.43301
3	10	9	2	0	0	1	4	0.00000	0.86603
4	11	10	3	0	0	1	4	−0.75000	0.43301
5	12	11	4	0	0	1	4	−0.75000	−0.43301
6	13	12	5	0	0	1	4	0.75000	−0.43301
1	8	13	6	0	0	1	4	0.00000	−0.86603

三边值问题)；为了在给定一个最简单的初始分划信息后，机器能自动剖分，并输出任意细密的网格剖分的各种信息，以便于形

成总刚阵、右端项且进行有限元方程的求解；可编制一个区域分划通用子程序。同时需约定对初始分划Ω_1的要求和对边界属性的要求。

对初始分划的要求。对于任意给定的区域，要求手工初始分划是**强正规的**(参见第四章§1)，即把区域分成大小相当的锐角或直角三角形，这一要求对于一般的有限元计算来说，可以免除，若要使用超收敛性(参见第四章§1)，此要求不能免除；要求各种边值条件的边界交点必须是初始分划节点；要求初始分划中某节点的邻节点必须按反时针方向排列于邻节点表中；对于内节点，其邻点表中的起始点和终点都没有要求；对于边界点，要求其邻点表中的起始点和终止点都是边界点，且按反时针方向排列。

对边界节点边界属性的约定。若节点 J 对应于第一边值节点、第二边值节点或第三边值节点，则LD(J)的数值分别为 0、1或 8 。

对区域自动剖分子程序的输入输出形式的要求。输入Ω_{k-1}网格层上的节点信息表，输出Ω_k网格层上的节点信息表。这样，只要输入一个信息量很少的初始分划Ω_1的节点信息表，并规定细分的次数 ($l-1$)，经机器自动分划后就可获得任意细密程度的网格Ω_l和该网格层上的节点信息表，这就为在Ω_l层上自动形成有限元方程提供一切必要的信息。

§4 有限元方程的形成

形成有限元方程的右端项是比较容易的，只需按第 2 章的计算公式(2.29)和(2.36)并调用数值积分子程序即可。当编号为 j 的节点是内点时，它所对应的有限元方程右端项元素只需计算Ⅲ(公式(2.29))；当 j 是第一边值点时，其对应右端项恒为该点处的边界值 u_{j}；当 j 是第二或第三边值点时，要按公式(2.29)和(2.36)计算Ⅲ + Ⅴ。最后还应注意，有限元方程的右端项还应包括总刚阵形成过程中的第一边值处理的结果。

形成总刚阵比较困难些，特别是在考虑到机器自动剖分、大型稀疏的总刚阵元素的压缩存贮和微分方程各种边值条件的处理等情况下，机器自动形成总刚阵就更复杂。尽管如此，人们还是可以利用其中存在的很强的规律，编制形成总刚阵的**通用程序**。

为此，首先总结在形成总刚阵时数值计算的特征，并将其列于表 4 中。表 4 中的 i 和 j 表示网格节点的编号，a_{ij}表示总刚阵元素且 $a_{ij}=a_{ji}$。

按表 4，图3.6～图3.10给出了总刚阵形成的 FORTRAN 流程框图，它适用于第一、三混合边值问题。图3.6÷3.10中各变量的意义和作用是：

LB	邻点标号(数组)，	LC	节点阶数(数组)，
LD	节点边界属性(数组)，	LK	总刚阵元素地址(数组)，
E	总刚阵元素存贮(数组)，	F	右端载荷量(数量)，
X	节点的x坐标(数组)，	Y	节点的y坐标(数组)，
LH	诸节点阶数的上界，	N	节点总数，
AIJ	总刚阵元素A_{ij}，	BIJ	过渡单元，

$G(X(I),\ Y(I))$ 第一边值函数值$g(x_i,\ y_i)$，

$(\mathrm{I})=\int_{\Delta}(p(x,y)\nabla\varphi_i\nabla\varphi_j+r(x,y)\varphi_i\varphi_j)dxdy$，同(2.27)式，

$(\mathrm{IV})=\int_{\Gamma_2^e}\sigma(s)\varphi_i\varphi_j ds$，同(2.35)式。

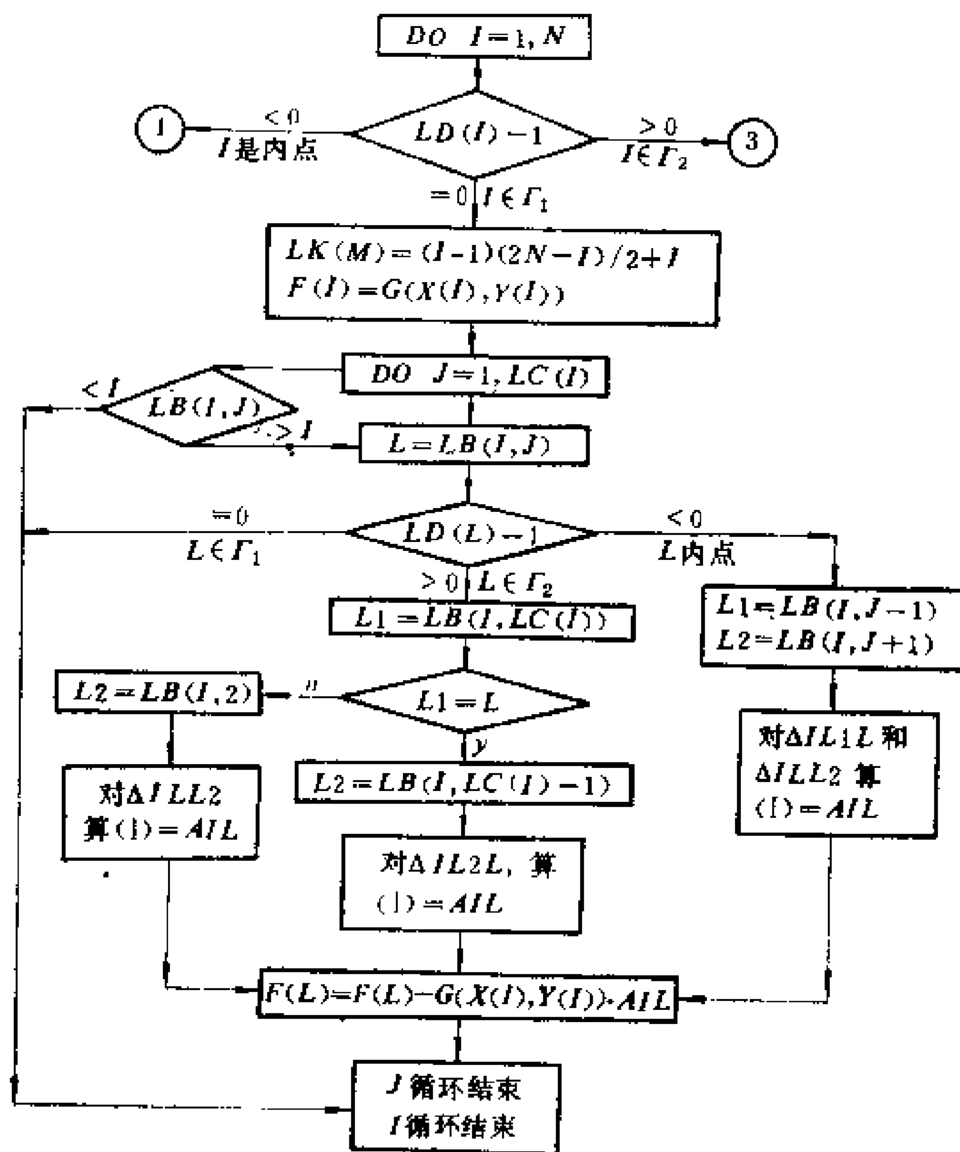
由于在有限元方法中选用了局部非零的基函数，同时由于$a(\cdot,\cdot)$的对称性，所以按表 4 规则形成的总刚阵是对称的且是大型稀疏的矩阵。这个总刚阵是对应于连同边界节点在内的整体节点编号情形，对应第一边值边界节点的编号，总刚阵的对角元素是 1，以这个编号的行和列中的其它元素全为零。若对此总刚阵元素采用变带宽压缩存贮法，则可用直接法解有限元方程；若采

用非零元素压缩存贮法，则只能用迭代法或基于迭代法基础的种种数值方法求解有限元方程组。

值得指出的是，对于常系数二阶椭圆型偏微分方程，当其定

表3.4 总刚阵元素计算规则

$a_{i,j}$（i 行，j 列）	内节点	第一边值点	第二、三边值点
内节点	$i\neq j$时，在以$\overline{ij}$为公共边的所有三角形上计算Ⅰ，得$a_{i,j}$；$i=j$时，在以i为顶点的所有三角形上计算Ⅱ，得$a_{i,i}$。	在以$\overline{ij}$为公共边的两个三角形上计算Ⅰ，并将Ⅰ的值乘以$(-u_j)$后加入右端项，而$a_{i,j}=0$。	在以$\overline{ij}$为公共边的两个三角形上计算Ⅰ，得$a_{i,j}$。
第一边值点	在以$\overline{ij}$为公共边的两个三角形上计算Ⅰ，并将Ⅰ的值乘以$(-u_i)$后加入右端项，而$a_{i,j}=0$。	$i\neq j$时，$a_{i,j}=0$；$i=j$时，$a_{i,i}=1$。	在以$\overline{ij}$为公共边的一个三角形上计算Ⅰ，并将Ⅰ的值乘以$(-u_i)$后加入右端项；而$a'_{i,j}=0$。
第二、三边值点	在以$\overline{ij}$为公共边的两个三角形上计算Ⅰ，得$a_{i,j}$。	在以$\overline{ij}$为公共边的一个三角形上计算Ⅰ，并将Ⅰ的值乘以$(-u_j)$后加入右端项，而$a_{i,j}=0$。	$i\neq j$时，在以$\overline{ij}$为公共边的一个三角形上计算Ⅰ＋Ⅳ，得$a_{i,j}$；$i=j$时，在以i为顶点的所有三角形上计算Ⅰ＋Ⅳ，得$a_{i,i}$。



DO I=1,N
LD(I)−1
<0
I是内点
>0
I∈Γ2
=0
I∈Γ1
LK(M)=(I-1)(2N−I)/2+I
F(I)=G(X(I),Y(I))
DO J=1,LC(I)
LB(I,J)
<I
>I
L=LB(I,J)
LD(L)−1
=0
L∈Γ1
<0
L内点
>0
L∈Γ2
L1=LB(I,LC(I))
L1=LB(I,J−1)
L2=LB(I,J+1)
L2=LB(I,2)
L1=L
n
y
L2=LB(I,LC(I)−1)
对ΔILL2 算(I)=AIL
对ΔIL2L，算(I)=AIL
对ΔIL1L和ΔILL2 算(I)=AIL
F(L)=F(L)−G(X(I),Y(I))·AIL
J循环结束
I循环结束

图3.6

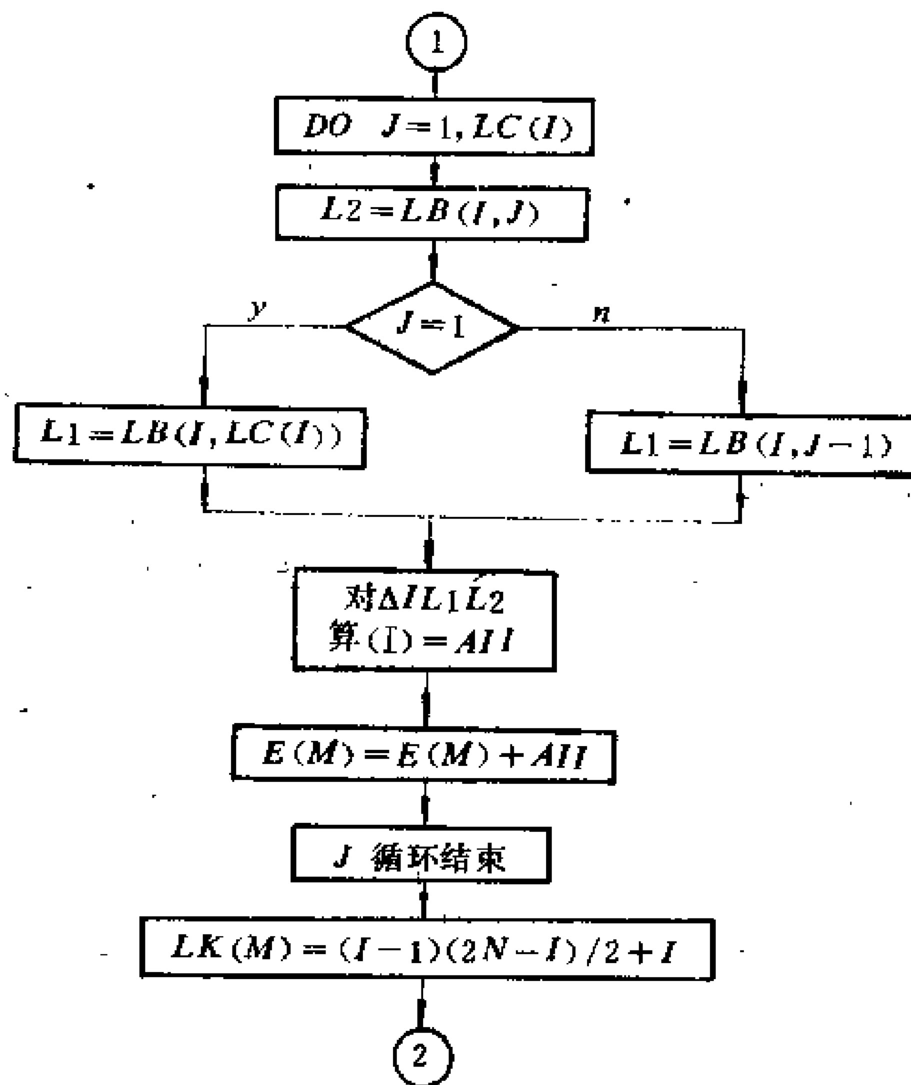
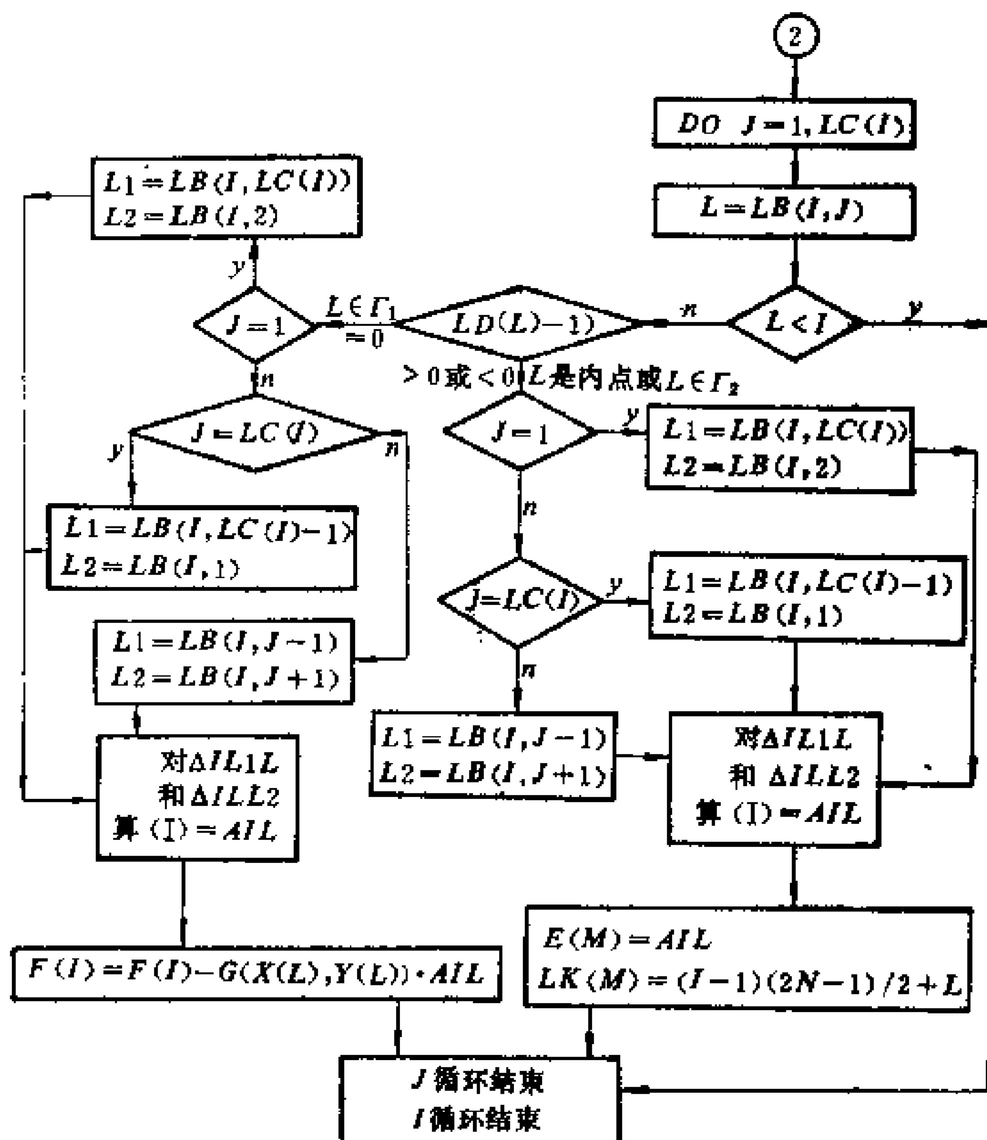
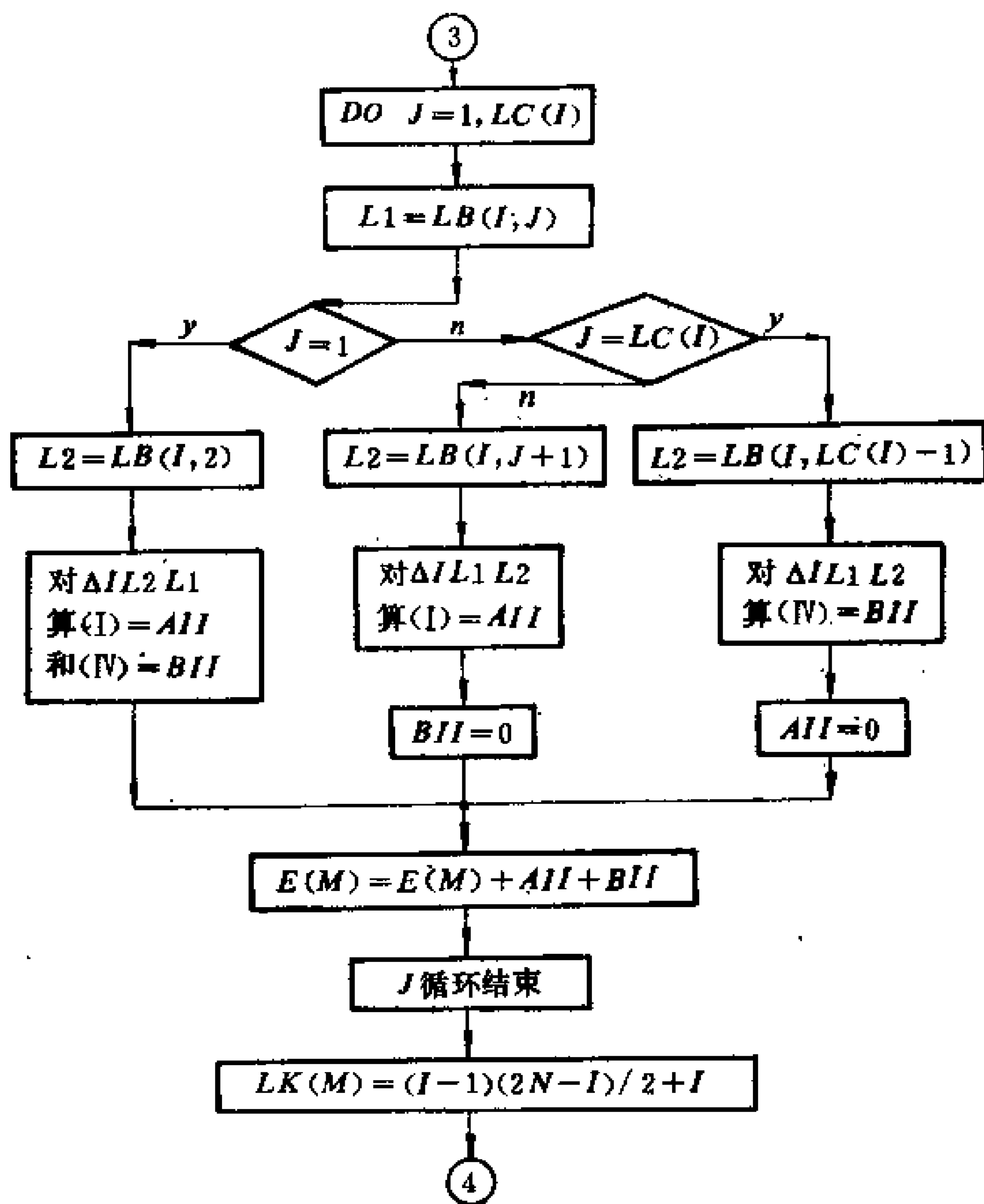

1
DO J=1,LC(I)
L2=LB(I,J)
J=1
y
n
L1=LB(I,LC(I))
L1=LB(I,J-1)
对ΔIL1L2
算(I)=AII
E(M)=E(M)+AII
J 循环结束
LK(M)=(I-1)(2N-I)/2+I
2

图3.7



2
DO J=1, LC(I)
L=LB(I,J)
L<I
y
n
LD(L)−1
L∈Γ1
=0
>0或<0
L是内点或L∈Γ2
J=1
y
n
L1=LB(I,LC(I))
L2=LB(I,2)
J=LC(I)
y
n
L1=LB(I,LC(I)−1)
L2=LB(I,1)
L1=LB(I,J−1)
L2=LB(I,J+1)
对ΔIL1L
和ΔILL2
算(I)=AIL
J=1
J=LC(I)
L1=LB(I,LC(I))
L2=LB(I,2)
L1=LB(I,LC(I)−1)
L2=LB(I,1)
L1=LB(I,J−1)
L2=LB(I,J+1)
对ΔIL1L
和ΔILL2
算(I)=AIL
F(I)=F(I)−G(X(L),Y(L))·AIL
E(M)=AIL
LK(M)=(I−1)(2N−1)/2+L
J循环结束
I循环结束

图3.8



3
DO J=1,LC(I)
L1=LB(I,J)
J=1
y
n
J=LC(I)
y
n
L2=LB(I,2)
L2=LB(I,J+1)
L2=LB(I,LC(I)-1)
对ΔIL2 L1
算(I)=AII
和(IV)=BII
对ΔIL1 L2
算(I)=AII
对ΔIL1 L2
算(IV)=BII
BII=0
AII=0
E(M)=E(M)+AII+BII
J循环结束
LK(M)=(I-1)(2N-I)/2+I
4

图3.9

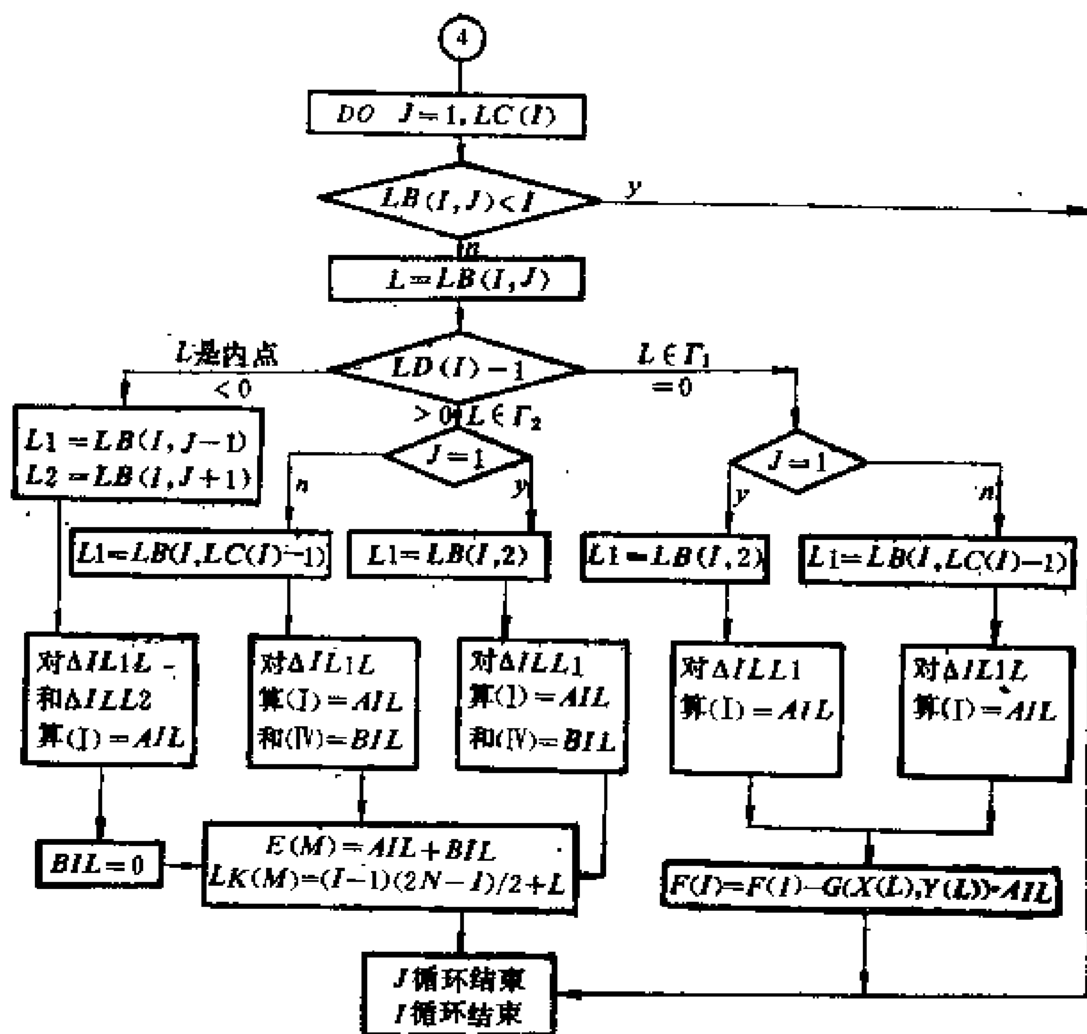

4
DO J=1,LC(I)
LB(I,J)<I
y
n
L=LB(I,J)
L是内点
<0
LD(I)-1
L∈Γ1
=0
>0 L∈Γ2
L1=LB(I,J-1)
L2=LB(I,J+1)
J=1
n
y
J=1
y
n
L1=LB(I,LC(I)-1)
L1=LB(I,2)
L1=LB(I,2)
L1=LB(I,LC(I)-1)
对ΔIL1L和ΔILL2算(I)=AIL
对ΔIL1L算(I)=AIL和(IV)=BIL
对ΔILL1算(I)=AIL和(IV)=BIL
对ΔILL1算(I)=AIL
对ΔIL1L算(I)=AIL
BIL=0
E(M)=AIL+BIL
LK(M)=(I-1)(2N-I)/2+L
F(I)=F(I)-G(X(L),Y(L))*AIL
J循环结束
I循环结束

图3 10

解区域为多边形(特别是矩形)且采用直角三角形分划时，不仅区域分划的节点编号非常规则，而且总刚阵元素都是常数，无需数值积分，总刚阵有时甚至是规则的 5 对角(或 7 对角)矩阵。所以，不必按前面介绍的通用程序形成有限元方程，只要在求解线代数方程组过程中(如用松弛迭代法)直接取用(2.30)或(2.32)式并注意边值处理即可。例如用有限元方法求解 poisson 方程；对矩形区域采用等腰直角三角形分划，对所有节点按行、列进行坐标编号，记为(i,j)；由于有限元方法离散和迭代过程的局部性；所以求解有限元方程时，只需对一切节点(i,j)循环，按(SOR 方法)公式

$$u_{ij}^{k+1}=(1-\omega)u_{ij}^{k}+\frac{\omega}{4}(u_{i,j+1}^{k}+u_{i+1,j}^{k}+u_{i-1,j}^{k+1}+u_{i,j-1}^{k+1})+\frac{\omega h^{2}}{4}f_{ij}$$

计算并注意边值处理，即可很方便地获得数值结果。

§5 有限元方法计算流程

本章前几节已详细叙述了使用有限元方法时必须采用和应该采用的各种程序技巧和特征。在此基础上读者不难想象，若输入初始网格分划 Ω_1 的信息表，并规定机器进一步细分的次数 $l-1$，则机器自动剖分并可输出所期望的细网格层 Ω_l 的各种信息；根据这些网格分划信息和有限元方程中的各种信息(如(2.1)式中的 p、r、f、σ、α、β)，机器执行由图3.6～图3.10所示过程就可自动地利用压缩存贮并形成有限元方程。这一过程当然要常常调用求函数值子程序和数值积分子程序；再根据压缩存贮的形式，决定选用某种可行的解线性代数方程组的算法；对于 Lagrange 型插值函数作为有限元近似函数情形，解有限元方程获得的是Ω_l网格分划节点(编号为 i)处的离散值 u_i，它是该节点处位移函数 u 的近似值。有限元位移值输出可设置于输出子程序中，输出子程序中

还可设置其它功能，例如，利用外推和超收敛性可输出高精度的位移值和导数值(参见第四章§1)，又例如误差估计或其它所需要的一切结果。

据此，有限元方法的计算流程可大致表述于流程图3.11中。

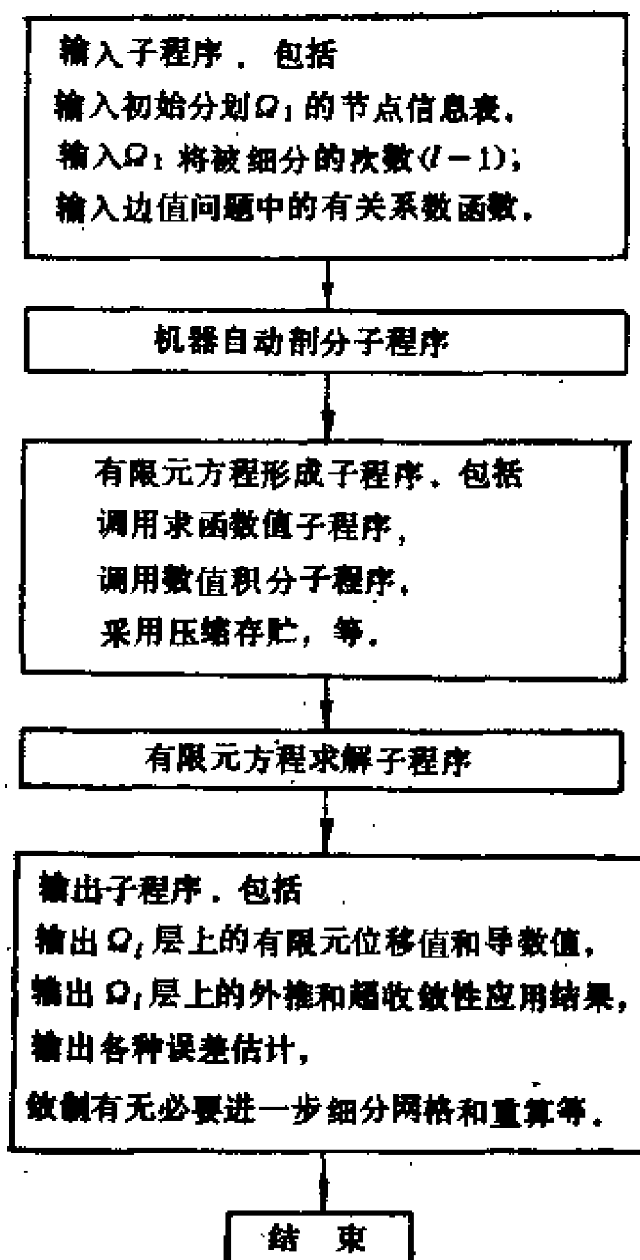


图3.11

§6　有限元方法在应用中的一些其它问题

6.1　间断系数问题

在实际问题中，常会遇到介质系数有间断的情形。例如，一个简单的**膜平衡问题**（如图3.12所示），Ω^-和Ω^+上张有两种不同属性的膜，l是其分界线，那么势函数应符合下面拟调和方程

$$\begin{cases} Lu=-\left(\dfrac{\partial}{\partial x}\left(\beta\dfrac{\partial u}{\partial x}\right)+\dfrac{\partial}{\partial y}\left(\beta\dfrac{\partial u}{\partial y}\right)\right)=f(x,y), \\ \qquad\qquad\qquad\qquad\qquad\qquad (x,y)\in\Omega/l; \\ \left(\beta\dfrac{\partial u}{\partial n}\right)^-+\left(\beta\dfrac{\partial u}{\partial n}\right)^+=0,\ (x,y)\in l; \end{cases}$$

$$\beta\frac{\partial u}{\partial n}+\eta u=g,\quad (x,y)\in l_1;$$

$$u=\bar{u},\qquad (x,y)\in l_2;$$

$$\beta(x,y)=\begin{cases}\beta_1(x,y), & (x,y)\in\Omega^-;\\ \beta_2(x,y), & (x,y)\in\Omega^+;\end{cases}$$

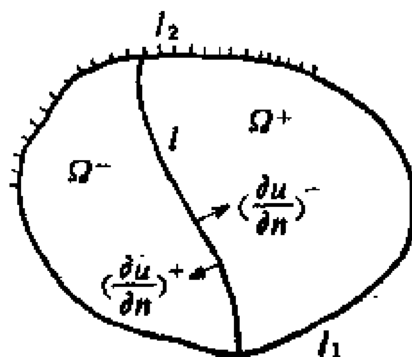


图3.12

其中，β_1和β_2分别是Ω^-和Ω^+上的介质系数；$\left(\beta\frac{\partial u}{\partial n}\right)^-$表示位于$\Omega^-$边界$l$的内侧并指向$\Omega^+$方向的法向导数，$\left(\beta\frac{\partial u}{\partial n}\right)^+$表示位于$\Omega^+$边界$l$的内侧并指向$\Omega^-$方向的**法向导数**，它们符合**广义自然边界条件**。

由于β在Ω上不连续，所以对上述方程直接应用Green公式将其转化为弱解方程是不允许的，但可以分别在Ω^-和Ω^+上将原方程化成相应的弱解方程。故由

$$\int_{\Omega^-}Luv\,dx\,dy=\int_{\Omega^-}\beta_1\nabla u\nabla v\,dxdy-\int_{\partial\Omega^-}\beta_1\frac{\partial u}{\partial n}v\,ds$$

$$=\int_{\Omega^-}\beta_1\nabla u\nabla v\,dxdy-\int_{\partial\Omega^-\cap l_1}\beta_1\frac{\partial u}{\partial n}v\,ds-\int_l\left(\beta_1\frac{\partial u}{\partial n}\right)^-v\,ds;$$

和

$$\int_{\Omega^+}Luv\,dx\,dy=\int_{\Omega^+}\beta_2\nabla u\nabla v\,dx\,dy-\int_{\partial\Omega^+}\beta_2\frac{\partial u}{\partial n}v\,ds$$

$$=\int_{\Omega^+}\beta_2\nabla u\nabla v\,dxdy-\int_{\partial\Omega^+\cap l_1}\beta_2\frac{\partial u}{\partial n}v\,ds-\int_l\left(\beta_2\frac{\partial u}{\partial n}\right)^+v\,ds$$

有

$$\int_\Omega Luv\,dx\,dy=\int_{\Omega^-\cup\Omega^+}\beta\nabla u\nabla v\,dxdy-\int_{l_1}\beta\frac{\partial u}{\partial n}v\,ds$$

$$-\int_l\left[\left(\beta_1\frac{\partial u}{\partial n}\right)^-+\left(\beta_2\frac{\partial u}{\partial n}\right)^+\right]v\,ds=\int_\Omega\beta\nabla u\nabla v\,dx\,dy$$

$$-\int_{l_2}\beta\frac{\partial u}{\partial n}v\,ds.$$

因此，不论有无间断系数分界线，与微分方程相应的弱解方程的形式是一样的。用有限元方法处理间断系数问题时，只需注意在间断线上安排节点，用折线近似间断线，并区别 Ω^-和 Ω^+区域上的间断系数即可，这是有限元方法的优点之一。

6.2 对称性处理

当定解区域、介质系数、载荷条件、边值条件等都有某种对称性时(如轴对称或中心对称等)，微分方程的解（及相应的数值解）也具有相应的对称性，在对称线上与解曲面相切且垂直于对称线方向解的导数等于零。利用这种对称性，可使求数值解的工作得以简化。例如，设定解区域如图3.13 a)所示，它关于 $\overline{a_3b}$ 对称；若定解问题

$$\begin{cases} -\Delta u=f(x,y), & (x,y)\in\Omega; \\ u=g, & (x,y)\in\overline{a_1a_2}; \\ \dfrac{\partial u}{\partial n}+\eta u=q, & (x,y)\in\overline{a_1a_3}\cup\overline{a_3a_2} \end{cases}$$

中的 f、g、η 和 q 都关于 $\overline{a_3b}$ 对称；则此定解问题的解也关于 $\overline{a_3b}$ 对称，且有$\left.\dfrac{\partial u}{\partial n}\right|_{\overline{a_3b}}=0$，这样问题便转化为

$$\begin{cases} -\Delta u=f(x,y), & (x,y)\in\triangle a_1ba_3; \\ u=g, & (x,y)\in\overline{a_1b}; \\ \dfrac{\partial u}{\partial n}+\eta u=q, & (x,y)\in\overline{a_1a_3}; \\ \dfrac{\partial u}{\partial n}=0, & (x,y)\in\overline{a_3b} \end{cases}$$

在部分区域上的求解问题。

同样可将此方法推广到有多个对称轴的情形和中心对称情形，见图 3.13b)、c)。例如，若解在图 3.13b)的区域上对称，

则只须在三角形 abc 上求解就可以了。

这种对称性处理可大大节省计算量和存贮量。

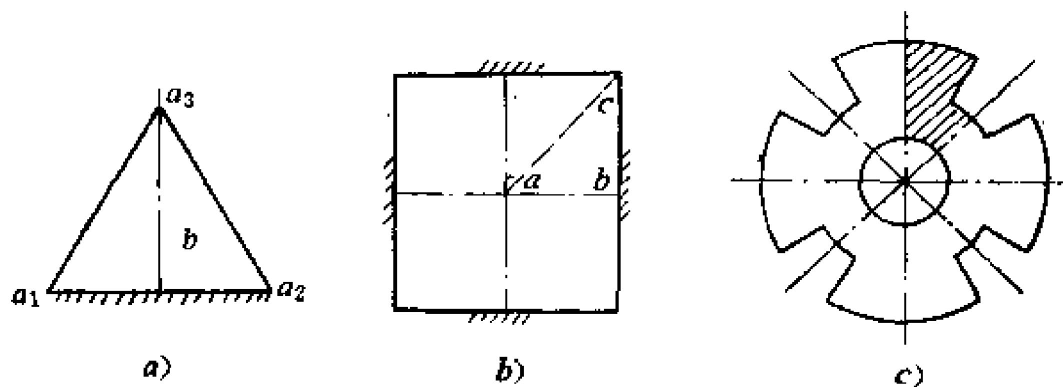


图3.13

6.3 三维问题简化为二维问题

水库的稳定渗流问题，可简单地取某个切面研究之，其数学模型是

$$\begin{cases} -k_1\dfrac{\partial^2 u}{\partial x^2}-k_2\dfrac{\partial^2 u}{\partial y^2}=0, & (x,y)\in\Omega; \\ k_1\dfrac{\partial u}{\partial x}n_x+k_2\dfrac{\partial u}{\partial y}n_y=0, & (x,y)\in BCDH\cup AER; \\ u=p/\rho+h, & (x,y)\in AB\cup HR; \end{cases}$$

其中 k_1、k_2 和 n_x、n_y 表示 x、y 方向的渗透系数和法向导数分量，p 是水压，ρ 是水的密度，h 是水面高，u 是渗流水势函数，参见图3.14。

确定在旋转体(图3.15)上的三维边值问题：

$$\begin{cases} -\left[\dfrac{\partial}{\partial x}\left(\beta\dfrac{\partial u}{\partial x}\right)+\dfrac{\partial}{\partial y}\left(\beta\dfrac{\partial u}{\partial y}\right)+\dfrac{\partial}{\partial z}\left(\beta\dfrac{\partial u}{\partial z}\right)\right]=f(x,y,z), \\ \qquad\qquad\qquad\qquad\qquad\qquad (x,y,z)\in\Omega; \\ \beta\dfrac{\partial u}{\partial n}+\eta u=q, \quad (x,y)\in S_1(\text{外表面}); \\ u=g, \qquad\qquad (x,y)\in S_2(\text{内表面}); \end{cases} \tag{3.3}$$

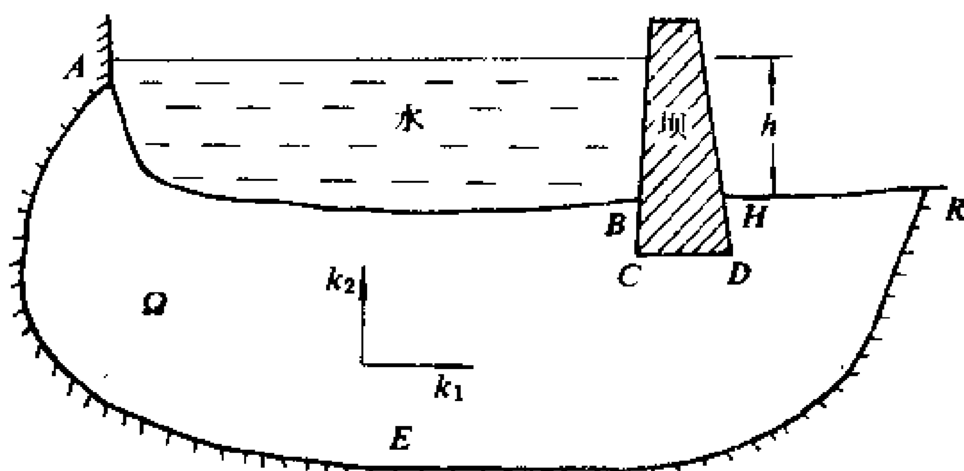


图3.14

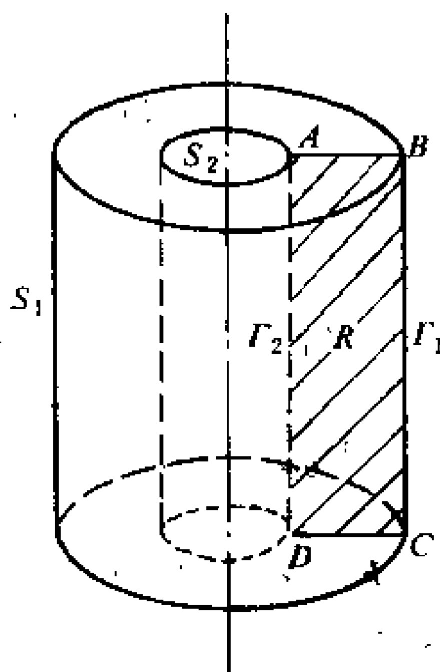


图3.15

其相应的变分问题是

$$J(u)=\int_{\Omega}\left\{\frac{\beta}{2}\left[\left(\frac{\partial u}{\partial x}\right)^2+\left(\frac{\partial u}{\partial y}\right)^2+\left(\frac{\partial u}{\partial z}\right)^2\right]-fu\right\}dxdydz+\int_{S_1}\left[\frac{1}{2}\eta u^2-qu\right]d\sigma\Rightarrow\min$$

$$(u|_{S_2}=g)。\quad(3.4)$$

若采用柱面坐标，就有

$$\left(\frac{\partial u}{\partial x}\right)^2+\left(\frac{\partial u}{\partial y}\right)^2+\left(\frac{\partial u}{\partial z}\right)^2=\left(\frac{\partial u}{\partial r}\right)^2+\left(\frac{\partial u}{\partial z}\right)^2,$$

$$\int_{\Omega}dxdydz=2\pi\int_R rdrdz,$$

$$\int_{S_i}d\sigma=2\pi\int_{\Gamma_i}rdz,\qquad i=1,2,$$

所以三维变分问题在采用柱面坐标后变为相应的二维变分问题：

$$J(u)=2\pi\int_R\left\{\frac{\beta}{2}\left[\left(\frac{\partial u}{\partial r}\right)^2+\left(\frac{\partial u}{\partial z}\right)^2\right]-fu\right\}rdrdz+2\pi\int_{\Gamma_1}\left(\frac{1}{2}\eta u^2-qu\right)rdz\Rightarrow\min\quad(u|_{\Gamma_2}=g)。\qquad(3.5)$$

这等价于二维边值问题：

$$
\begin{cases}
-\left[\dfrac{1}{r}\dfrac{\partial}{\partial r}\left(r\beta\dfrac{\partial u}{\partial r}\right)+\dfrac{\partial}{\partial z}\left(\beta\dfrac{\partial u}{\partial z}\right)\right]=f, & (r,z)\in R; \\
\beta\dfrac{\partial u}{\partial n}+\eta u=q, & (r,z)\in \Gamma_1; \\
u=g, & (r,z)\in \Gamma_2.
\end{cases}
\tag{3.6}
$$

应该注意，采用柱面坐标将三维边值问题(3.3)化为二维边值问题(3.6)后，在 $r=0$ 处解出现奇异，这是与事实不符的；但是(3.6)式等价的变分形式(3.5)式又恢复了问题本来的特征，弱解不再出现奇异点了，这也体现了有限元方法利用变分形式作为出发点的优越性。

其次应该看到，不论是在直角坐标系中求解(3.4)式，还是在柱面坐标系中求解(3.5)式，所采用的有限元方法的步骤是完全一样的，所以有限元方法是不受坐标系限制的。

最后应注意，若能将三维问题简化为二维问题，则能极大地节省计算量和存贮量。当然也有些问题并不能化为二维问题，那就必须将三维区域划分为若干个三维单元，再确定三维的节点基函数，并形成和求解有限元方程，尽管此时有限元方法的基本思想和实现步骤仍然同于二维问题，但程序设计要比二维情形复杂得多，由于篇幅限制，本书不能介绍解三维椭圆边值问题的有限元方法，有兴趣的读者可参阅[13]、[14]。

第四章　提高二维有限元解精度的讨论

提高有限元解精度的问题一直是值得深入研究的课题，本章只简单介绍有关采用高次元[8~15]、超收敛性[16,22]和外推[20~23]方面的最新成果及其应用，至于高效率高精度求解有限元方程的算法将在第三篇中讨论。

§1　三角线性元解的超收敛性和外推

利用三角线性元形成的有限元方程，解之则得解函数 $u(x,y)$ 在各个节点处的近似值 $(u_1,u_2,\cdots,u_N)$，有限元解可表示为

$$u_h(x,y)=\sum_{i=1}^{N}u_i\varphi(x,y).$$

利用公式(2.15)还可以求得任一单元△上 $\frac{\partial u}{\partial x}$ 和 $\frac{\partial u}{\partial y}$ 的近似值：

$$\left.\frac{\partial u_h}{\partial x}\right|_{\Delta}=\frac{1}{2S_{\Delta}}[u_i(y_j-y_k)+u_j(y_k-y_i)+u_k(y_i-y_j)],$$

$$\left.\frac{\partial u_h}{\partial y}\right|_{\Delta}=\frac{1}{2S_{\Delta}}[u_i(x_k-x_j)+u_j(x_i-x_k)+u_k(x_j-x_i)],$$

式中 S_{Δ} 表示单元△的面积。在某个单元上有限元解及其导数的精度是

$$\|u-u_h\|_0=\left(\int_{\Delta}|u-u_h|^2dxdy\right)^{1/2}=ch^2\|u\|_2,$$

$$\|u-u_h\|_1=\left(\sum_{|i|=0}^{1}\|D^i(u-u_h)\|_0^2\right)^{1/2}=ch\|u\|_2,$$

$$|u-u_h|_{\infty}=\max_{P\in\Delta}|u(P)-u_h(P)|=ch^2|\ln h|\,\|u\|_{2,\infty},$$

$$\left|\frac{\partial u}{\partial x}-\frac{\partial u_h}{\partial x}\right|_{\infty}=\max_{P\in\Delta}\left|\frac{\partial u(P)}{\partial x}-\frac{\partial u_h(P)}{\partial x}\right|=ch|\ln h|\,\|u\|_{2,\infty},$$

$$\left|\frac{\partial u}{\partial y}-\frac{\partial u_h}{\partial y}\right|_\infty=\max_{P\in\Delta}\left|\frac{\partial u(P)}{\partial y}-\frac{\partial u_h(P)}{\partial y}\right|=ch|\ln h|\|u\|_{2,\infty}.$$

上述这些误差估计的证明可参见第五章。

为了在所有节点上都获得较好的数值解精度，除要求三角剖分满足第二章 §2.1 中所述的一般规则外，还要求剖分是**强正规**的，即任两个相邻三角形单元构成 h^2-近似平行四边形，也就是要求它们所构成的四边形的对边长度差是 $O(h^2)$ 量级，其内错角之差为 $O(h)$ 量级。

定理 4.1[20~22] 设二阶线性椭圆边值问题的解 $u\in H^3(\Omega)$，凸区域 Ω 的分划是强正规的，则存在一个与 h 无关的函数 $w(z)$，使线性有限元解适合**渐近展式**：

$$u_h(z)=u^I(z)+w(z)h^2+O(h^3); \tag{4.1}$$

若是正三角形分划且 $u\in C^5(\Omega)$，则有渐近展式：

$$u_h(z)=u^I(z)+w(z)h^4+O(h^5), \tag{4.2}$$

式中 u_h 和 u^I 分别表示有限元解和 u 的插值函数；z 是 Ω_h 中的任意点。

利用有限元解的渐近展式 (4.1) 或 (4.2) 式即可施行外推技巧，从而提高有限元位移值的精度。

定理 4.2 设二阶线性椭圆边值问题的解 $u\in H^3(\Omega)$，凸区域 Ω 的分划是强正规的；将 Ω_{2h} 各单元边中点连接得分划 Ω_h；且记 u_{2h} 和 u_h 分别是剖分 Ω_{2h} 和 Ω_h 的有限元解；设 M_0、$M_0'\in\Omega_{2h}$，M 是 Ω_{2h} 分划中任一三角单元的边中点，即 $M\in\Omega_h$；那么第一边值问题三角线性元有限元解适合下述外推公式

$$\left|\frac{4u_h(M_0)-u_{2h}(M_0)}{3}-u(M_0)\right|_\infty=O(h^3), \tag{4.3}$$

$$\left|u_h(M)+\frac{1}{6}[u_h(M_0)-u_{2h}(M_0)]+\frac{1}{6}[u_h(M_0')-u_{2h}(M_0')]-u(M)\right|_\infty$$
$$=O(h^3). \tag{4.4}$$

证明 如图4.1所示，因为$M_0 \in \Omega_{2h} \cap \Omega_h$ 且 $u_h^I(M_0) = u_{2h}^I(M_0) = u(M_0)$，所以由渐近展式(4.1)得

$$u_{2h}(M_0) = u(M_0) + w(M_0)(2h)^2 + O(h^3),$$

$$u_h(M_0) = u(M_0) + w(M_0)h^2 + O(h^3),$$

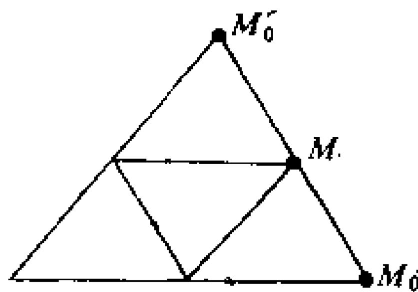


图 4.1

故有

$$\left| \frac{1}{3}[4u_h(M_0) - u_{2h}(M_0)] - u(M_0) \right|_\infty = O(h^3),$$

或

$$u(M_0) = \frac{1}{3}[4u_h(M_0) - u_{2h}(M_0)] + O(h^3).$$

又因 $M \in \Omega_h$，$M \notin \Omega_{2h}$，且

$$u_{2h}^I(M) = \frac{1}{2}[u(M_0) + u(M_0')],$$

$$u_{2h}(M) = \frac{1}{2}[u_{2h}(M_0) + u_{2h}(M_0')],$$

所以由(4.1)式又有

$$u_{2h}(M) = u_{2h}^I(M) + w(M)(2h)^2 + O(h^3),$$
$$u_h(M) = u(M) + w(M)(h^2) + O(h^3),$$

由此得到

$$u(M) = u_h(M) - w(M)h^2 + O(h^3)$$

$$= u_h(M) - \frac{1}{4}[u_{2h}(M) - u_{2h}^I(M)] + O(h^3)$$

$$= u_h(M) - \frac{1}{8}[u_{2h}(M_0) + u_{2h}(M_0')]$$

$$+ \frac{1}{8}[u(M_0) + u(M_0')] + O(h^3)$$

$$= u_h(M) - \frac{1}{8}[u_{2h}(M_0) + u_{2h}(M'_0)]$$

$$+ \frac{1}{24}[4u_h(M_0) - u_{2h}(M_0)]$$

$$+ \frac{1}{24}[4u_h(M'_0) - u_{2h}(M'_0)] + O(h^3)$$

$$= u_h(M) + \frac{1}{6}[u_h(M_0) - u_{2h}(M_0)]$$

$$+ \frac{1}{6}[u_h(M'_0) - u_{2h}(M'_0)] + O(h^3).$$

所以，证得外推公式(4.3)式和(4.4)式成立。 □

定理4.3 若二阶线性椭圆边值问题的解 $u \in C^5(\Omega)$，且凸区域 Ω 采用正三角形分划，那么 Ω_h 网格节点上的线性元解适合下述外推关系：

$$\left|\frac{1}{15}[16u_h(M_0) - u_{2h}(M_0)] - u(M_0)\right|_\infty = O(h^5), \qquad (4.5)$$

$$\left|u_h(M) + \frac{1}{30}[u_h(M_0) - u_{2h}(M_0)] + \frac{1}{30}[u_h(M'_0) - u_{2h}(M'_0)]\right.$$

$$\left. - u(M)\right|_\infty = O(h^5). \qquad (4.6)$$

利用渐近展式(4.2)且仿照定理4.2的证明，定理 4.3 是不难证明的。

人们只要算出 Ω_{2h} 和 Ω_h 网格层上的线性有限元解，即可利用外推公式(4.3)和(4.4)或(4.5)和(4.6)得到 Ω_h 上一切网格节点的高精度位移值。

对于三角线性元导数值，$\partial u_h/\partial x$ 和 $\partial u_h/\partial y$（常数）在某单元任一点处一般只有 $O(h)$ 的精度，但利用有限元解超收敛性研究的结果，也可以获得某些点处的高精度的导数值。

取相邻两单元 e_1 和 e_2 导数值的平均值

$$
\begin{aligned}
\bar{u}_x &= \frac{1}{2}\left[\left(\frac{\partial u_h}{\partial x}\right)_{e_1}+\left(\frac{\partial u_h}{\partial x}\right)_{e_2}\right],\\
\bar{u}_y &= \frac{1}{2}\left[\left(\frac{\partial u}{\partial y}\right)_{e_1}+\left(\frac{\partial u_h}{\partial y}\right)_{e_2}\right],
\end{aligned}
\tag{4.7}
$$

作为导数 u_x 和 u_y 在公共边中点 G（见图 4.2a)）处的导数近似值时，已经证明[14]它有二阶精度，即

$$|u_x(G)-\bar{u}_x|=O(h^2),\qquad |u_y(G)-\bar{u}_y|=O(h^2).$$

类似地，对于Ω_h的内点G，如图 4.2b)～d)所示，可分别利用相应于图中的 2、4、6 个单元的导数值作平均，用这个平均值作为G点的导数近似值也具有 $O(h^2)$ 的精度。

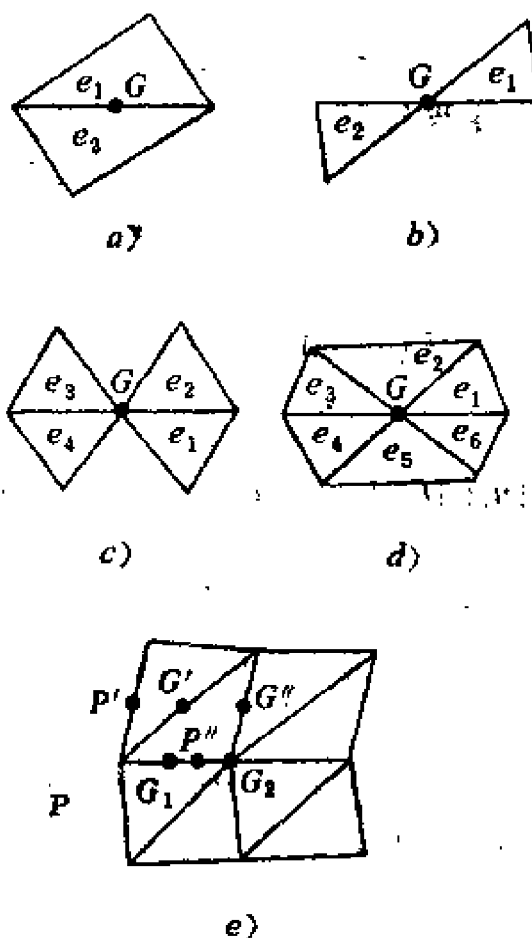


图 4.2

实际问题中，人们也希望得到其它点(特别是边界点)处的高精度的导数近似值。这时，也可利用相邻 2 个高精度导数近似值进行线性内插或作外推，所获得的导数近似值仍具有$O(h^2)$的精度。例如图4.2e)的剖分满足强正规条件，G_1是$\overline{PG_2}$的中点，G'和 P'' 分别是$\overline{P'G''}$和$\overline{G_1G_2}$的中点，利用G_1、G_2、G' 和 G'' 处的高精度导数值作线性内插和线性外推：

$$
\begin{aligned}
&\left(\frac{\partial u_h}{\partial x}\right)_{P''} = \frac{1}{2}[(\bar{u}_x)_{G_1} + (\bar{u}_x)_{G_2}],\\
&\left(\frac{\partial u_h}{\partial y}\right)_{P''} = \frac{1}{2}[(\bar{u}_y)_{G_1} + (\bar{u}_y)_{G_2}],\\
&\left(\frac{\partial u_h}{\partial x}\right)_{P} = \frac{1}{2}[3(\bar{u}_x)_{G_1} - (\bar{u}_x)_{G_2}],\\
&\left(\frac{\partial u_h}{\partial y}\right)_{P} = \frac{1}{2}[3(\bar{u}_y)_{G_1} - (\bar{u}_y)_{G_2}],\\
&\left(\frac{\partial u_h}{\partial x}\right)_{P'} = \frac{1}{2}[3(\bar{u}_x)_{G'} - (\bar{u}_x)_{G''}],\\
&\left(\frac{\partial u_h}{\partial y}\right)_{P'} = \frac{1}{2}[3(\bar{u}_y)_{G'} - (\bar{u}_y)_{G''}].
\end{aligned}
\tag{4.8}
$$

这样便可获得 P''、P 和 P' 处的具有 $O(h^2)$ 精度的导数近似值。

下面举一个简单的例子。在正方形区域 $\Omega=[-1,1]\times[-1,1]$上考虑边值问题

$$
\begin{cases}
-\Delta u = f(x,y) = x^2+y^2-2, & (x,y)\in\Omega,\\
u|_{\partial\Omega} = 0,
\end{cases}
$$

其准确解是 $u=-(1-x^2)(1-y^2)/2$。

首先将$[-1,1]\times[-1,1]$剖分成32个等面积的直角三角形单元，得分划 $\Omega_{2h}(2h=0.5\sqrt{2})$；再在 Ω_{2h} 基础上细分一次，即分成128 个等面积的直角三角形单元，得 $\Omega_h(h=0.25\sqrt{2})$；在 Ω_{2h} 和 Ω_h 上分别使用三角线性元；利用公式(4.5)和(4.6)求得相应节点处的高精度位移近似值 u_h^*；利用公式(4.7)和(4.8)求得相应点处的高精度导数近似值 $\bar{u}_x$ 和 $\bar{u}_y$。由于边值问题的右端项、边值条件和定解区域都关于 x、y 轴和 $x=y$ 直线对称，边值问题的解也关于这些对称轴对称，所以将 Ω 的 1/8 区域示于图 4.3 中；且将所求得的高精度数值解与真解的比较列于表 5 、表 6 和表 7 中。从表 6 和表 7 中的数值比较可见，本书中介绍的提高数值解精度的方法是有显著效果的。

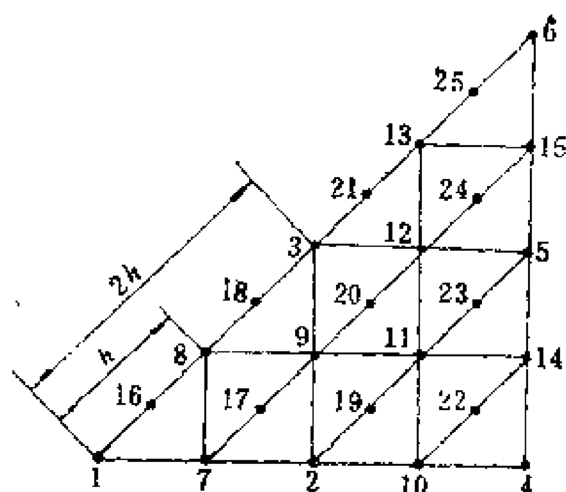


图 4.3

表 4.1　有限元解 u_{2h}（剖分 Ω_{2h}，$2h=0.71$）

节点号	准确解 u	有限元解 u_{2h}	精度 $c\times(2h)^2$ 中的系数 c
1	−0.5000	−0.5104	0.0208
2	−0.3750	−0.3553	0.0400
3	−0.2813	−0.2645	0.0336

表 4.2　有限元解 u_h 和外推解 u_h^*（剖分 Ω_h，$h=0.35$）

节点号	准确解 u	有限元解 u_h	精度 ch^2 中的系数 c	外推解 u_h^*	精度 ch^3 中的系数 c
1	−0.5000	−0.5054	0.0441	−0.5037	0.0862
2	−0.3750	−0.3676	0.0605	−0.3717	0.0769
3	−0.2813	−0.2755	0.0474	−0.2792	0.0489
7	−0.4688	−0.4644	0.0360	−0.4656	0.0746
8	−0.4395	−0.4335	0.0489	−0.4364	0.0723
9	−0.3516	−0.3449	0.0547	−0.3486	0.0699
10	−0.2188	−0.2122	0.0547	−0.2153	0.0815
11	−0.2051	−0.1991	0.0489	−0.2018	0.0769
12	−0.1641	−0.1601	0.0327	−0.1618	0.0536
13	−0.0957	−0.0933	0.0171	−0.0950	0.0163

表 4.3 高精度有限元导数值 $\overline{u_x}$ 和 $\overline{u_y}$

节点号	准确值 u_x 和 u_y		有限元佳值 $\overline{u_x}$ 和 $\overline{u_y}$		精度 $c\times0.35^2$ 中的系数 c	
16	0.1437	0.1437	0.1231	0.1231	0.1648	0.1648
17	0.3708	0.1073	0.3691	0.1074	0.0136	0.0008
18	0.3160	0.3160	0.3223	0.3223	0.0512	0.0512
19	0.6024	0.0716	0.6152	0.0762	0.1024	0.0368
20	0.5223	0.2168	0.5371	0.2285	0.1184	0.0936
21	0.3644	0.3644	0.3809	0.3809	0.1320	0.1320
22	0.8226	0.0261	0.8612	0.0293	0.3088	0.0256
23	0.7184	0.0780	0.7520	0.0879	0.2688	0.0792
24	0.5068	0.1336	0.5332	0.1465	0.2112	0.1032
25	0.1886	0.1886	0.2051	0.2051	0.1480	0.1480

§2 提高四边形双线性元解精度的讨论

2.1 矩形双线性元

除三角剖分外，四边形剖分也是常用的。较简单的情形是定解区域可分割为若干直边四边形的和，参见图 4.4 *a*)、*b*)、*c*)。

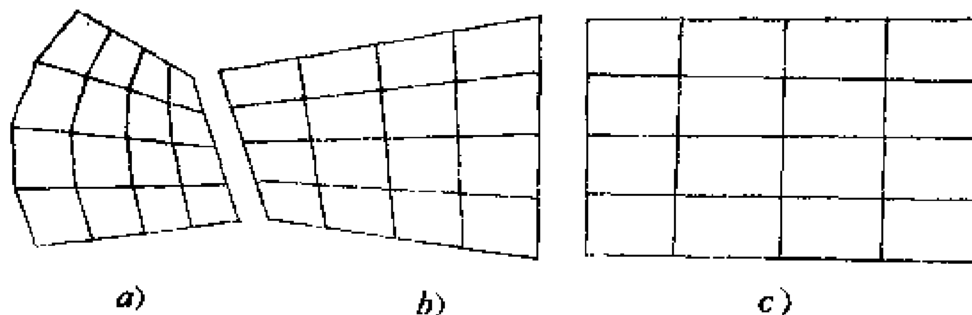


图 4.4

利用变换，可将一个凸直边四边形映为单位正方形(图4.5)，其变换为

$$\begin{cases} x = x_1 + (x_2 - x_1)\xi + (x_3 - x_1)\eta + (x_4 - x_2 - x_3 + x_1)\xi\eta, \\ y = y_1 + (y_2 - y_1)\xi + (y_3 - y_1)\eta + (y_4 - y_2 - y_3 + y_1)\xi\eta. \end{cases} \tag{4.9}$$

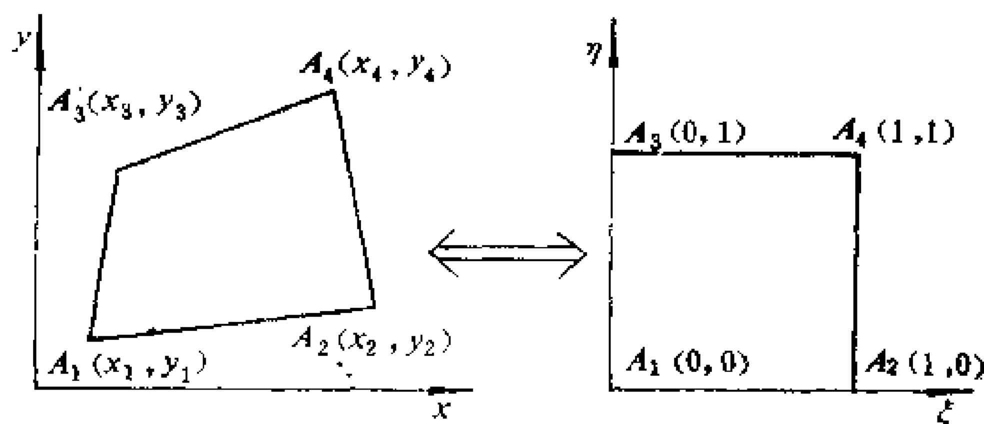


图4.5

因此，只要在单位正方形上构造出插值型函数，利用变换(4.9)就可得到任一个四边形单元上的插值函数，从而可构造出适合有限元方程需要的允许函数和试探函数了。

若在单位正方形E的4个顶点上给定节点值$u_i(i=1,2,3,4)$，那么在E上最简单的插值函数u_{hE}是**双线性函数**，且可表示为

$$u_{hE} = \sum_{i=1}^{4} u_i N_i(\xi, \eta),$$

式中$N_i(\xi,\eta)$为相应的单元节点基函数，它们都是双线性函数。所谓$N_i(\xi,\eta)$双线性，是指当ξ固定时，$N_i(\xi,\eta)$为关于η的线性函数；而当η固定时，$N_i(\xi,\eta)$又是关于ξ的线性函数。利用$N_i(\xi,\eta)$的双线性特点和它在相应节点处的表现，不难确定

$$\begin{cases} N_1(\xi,\eta) = (1-\xi)(1-\eta), & N_2(\xi,\eta) = \xi(1-\eta), \\ N_3(\xi,\eta) = \xi\eta, & N_4(\xi,\eta) = \eta(1-\xi), \end{cases} \tag{4.10}$$

单元节点基函数$N_i(\xi,\eta)$的图形见图4.6。

对于四边形分划Ω_h中任一节点P_i，整体节点基函数$\varphi_i(x,y)$是由P_i支集(以P_i为顶点的所有四边形)中各个单元上的基函数

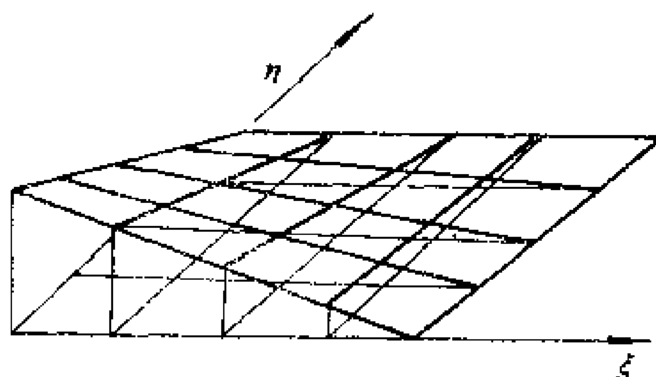


图 4.6

拼凑而成，而每个单元上相应于 P_i 点处的基函数可利用 (4.10) 式和(4.9)式得到。因此，在 Ω_h 上的有限元形函数

$$u_h(x, y) = \sum_{i=1}^{N} u_i \varphi_i(x, y)$$

是整体连续和局部(在各个单元上)双线性的。再利用整体节点基函数 $\varphi_i(x, y)$，则可形成有限元方程

$$\sum_{i=1}^{N} a(\varphi_i, \varphi_j) u_i = F(\varphi_j), \quad j = 1, 2, \cdots, N. \qquad (4.11)$$

特别地，设区域采用矩形分划，h_1 和 h_2 分别表示 x 和 y 方向的步长，节点(x_i, y_j)记为(i, j)，该节点处的整体节点基函数记为 φ_{ij}，以该节点为顶点的四个矩形单元记为 $E_{i\pm1, j\pm1}$，见图 4.7，那么整体节点基函数 φ_{ij} 可表示为

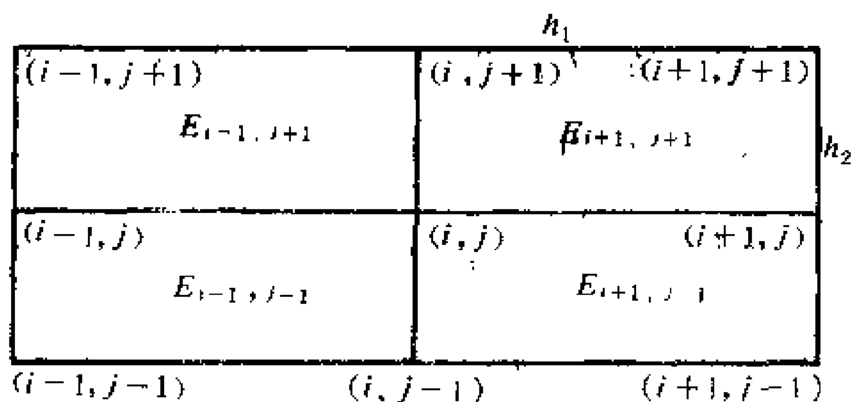


图 4.7

$$\varphi_{ij}=\begin{cases}\left(1-\left|\dfrac{x-x_i}{h_1}\right|\right)\left(1-\left|\dfrac{y-y_j}{h_2}\right|\right), & (x,y)\in E_{i\pm1,j\pm1},\\ 0, \quad (x,y)\text{在别处。}\end{cases}$$

再将坐标原点移至(x_i, y_j)，则在 $E_{i\pm1,j\pm1}$ 上9个整体节点基函数的限制为

$$\begin{cases}
\varphi_{ij}=\begin{cases}\left(1-\dfrac{\xi}{h_1}\right)\left(1-\dfrac{\eta}{h_2}\right), & (\xi,\eta)\in E_{i+1,j+1},\\ \left(1-\dfrac{\xi}{h_1}\right)\left(1+\dfrac{\eta}{h_2}\right), & (\xi,\eta)\in E_{i+1,j-1},\\ \left(1+\dfrac{\xi}{h_1}\right)\left(1-\dfrac{\eta}{h_2}\right), & (\xi,\eta)\in E_{i-1,j+1},\\ \left(1+\dfrac{\xi}{h_1}\right)\left(1+\dfrac{\eta}{h_2}\right), & (\xi,\eta)\in E_{i-1,j-1};\end{cases}\\
\varphi_{i+1,j}=\begin{cases}\dfrac{\xi}{h_1}\left(1-\dfrac{\eta}{h_2}\right), & (\xi,\eta)\in E_{i+1,j+1},\\ \dfrac{\xi}{h_1}\left(1+\dfrac{\eta}{h_2}\right), & (\xi,\eta)\in E_{i+1,j-1};\end{cases}\\
\varphi_{i-1,j}=\begin{cases}\dfrac{-\xi}{h_1}\left(1-\dfrac{\eta}{h_2}\right), & (\xi,\eta)\in E_{i-1,j+1},\\ \dfrac{-\xi}{h_1}\left(1+\dfrac{\eta}{h_2}\right), & (\xi,\eta)\in E_{i-1,j-1};\end{cases}\\
\varphi_{i,j+1}=\begin{cases}\left(1-\dfrac{\xi}{h_1}\right)\dfrac{\eta}{h_2}, & (\xi,\eta)\in E_{i+1,j+1},\\ \left(1+\dfrac{\xi}{h_1}\right)\dfrac{\eta}{h_2}, & (\xi,\eta)\in E_{i-1,j+1};\end{cases}\\
\varphi_{i,j-1}=\begin{cases}\left(1-\dfrac{\xi}{h_1}\right)\left(-\dfrac{\eta}{h_2}\right), & (\xi,\eta)\in E_{i+1,j-1},\\ \left(1+\dfrac{\xi}{h_1}\right)\left(-\dfrac{\eta}{h_2}\right), & (\xi,\eta)\in E_{i-1,j-1};\end{cases}\\
\varphi_{i+1,j+1}=\dfrac{\xi}{h_1}\cdot\dfrac{\eta}{h_2}, \qquad (\xi,\eta)\in E_{i+1,j+1};\\
\varphi_{i-1,j+1}=\dfrac{-\xi}{h_1}\cdot\dfrac{\eta}{h_2}, \qquad (\xi,\eta)\in E_{i-1,j+1};
\end{cases} \tag{4.12}$$

$$\varphi_{i-1,j-1}=\frac{\xi}{h_1}\cdot\frac{\eta}{h_2},\qquad (\xi,\eta)\in E_{i-1,j-1};$$

$$\varphi_{i+1,j-1}=\frac{-\xi}{h_1}\cdot\frac{\eta}{h_2},\qquad (\xi,\eta)\in E_{i+1,j-1}.$$

从而计算出解 **Poisson** 问题有限元方程总刚阵中与节点 (i,j) 有关的一行中的 9 个元素为

$$\begin{cases}
a(\varphi_{ij},\varphi_{i-1,j+1})=-\dfrac{1}{6}\left(\dfrac{h_2}{h_1}+\dfrac{h_1}{h_2}\right),\\
a(\varphi_{ij},\varphi_{i,j+1})=\dfrac{1}{3}\dfrac{h_2}{h_1}-\dfrac{2}{3}\dfrac{h_1}{h_2},\\
a(\varphi_{ij},\varphi_{i+1,j+1})=-\dfrac{1}{6}\left(\dfrac{h_2}{h_1}+\dfrac{h_1}{h_2}\right),\\
a(\varphi_{ij},\varphi_{i-1,j})=\dfrac{1}{3}\dfrac{h_1}{h_2}-\dfrac{2}{3}\dfrac{h_2}{h_1},\\
a(\varphi_{ij},\varphi_{ij})=\dfrac{4}{3}\left(\dfrac{h_2}{h_1}+\dfrac{h_1}{h_2}\right),\\
a(\varphi_{ij},\varphi_{i+1,j})=\dfrac{1}{3}\dfrac{h_1}{h_2}-\dfrac{2}{3}\dfrac{h_2}{h_1},\\
a(\varphi_{ij},\varphi_{i-1,j-1})=-\dfrac{1}{6}\left(\dfrac{h_2}{h_1}+\dfrac{h_1}{h_2}\right),\\
a(\varphi_{ij},\varphi_{i,j-1})=\dfrac{1}{3}\dfrac{h_2}{h_1}-\dfrac{2}{3}\dfrac{h_1}{h_2},\\
a(\varphi_{ij},\varphi_{i+1,j-1})=-\dfrac{1}{6}\left(\dfrac{h_2}{h_1}+\dfrac{h_1}{h_2}\right).
\end{cases}\tag{4.13}$$

若 $h_1=h_2$，则有

$$\begin{cases}
a(\varphi_{ij},\varphi_{i-1,j+1})=-\dfrac{1}{3}, & a(\varphi_{ij},\varphi_{i,j+1})=-\dfrac{1}{3},\\
a(\varphi_{ij},\varphi_{i+1,j+1})=-\dfrac{1}{3}, & a(\varphi_{ij},\varphi_{i-1,j})=-\dfrac{1}{3},\\
a(\varphi_{ij},\varphi_{ij})=\dfrac{8}{3}, & a(\varphi_{ij},\varphi_{i+1,j})=-\dfrac{1}{3},\\
a(\varphi_{ij},\varphi_{i-1,j-1})=-\dfrac{1}{3}, & a(\varphi_{ij},\varphi_{i,j-1})=-\dfrac{1}{3},\\
a(\varphi_{ij},\varphi_{i+1,j-1})=-\dfrac{1}{3}.
\end{cases}\tag{4.14}$$

2.2 关于提高四边形双线性元解精度的讨论

采用四边形线性元时，总刚阵元素的计算、非零元素的压缩存贮方法和第一边值的处理过程等都类似于三角元的有关讨论。求解有限元方程组，就得到有限元近似解 u_h 及各个单元上的有限元导数值 $\partial u_h/\partial x$ 和 $\partial u_h/\partial y$。

若剖分是正规的，有限元方程中各数值积分可采用二阶精度的数值积分公式(即采用类似于一维数值积分情形的中矩形或梯形积分公式)，也可采用具有三阶精度的数值积分公式（例如对 $[-1,1]\times[-1,1]$单元，选用与(0.57735, 0.7735)对称分布的4个高斯积分点的求积公式)，那么在单元节点处有限元解的精度一般为 $O(h^2)$ 或稍好一点，$\partial u_h/\partial x$ 和 $\partial u_h/\partial y$ 在任一单元上一般只有 $O(h)$ 精度，沒有更好的结果。

若剖分是强正规的(即除了满足正规剖分条件外，还要求每个单元是 h^2 阶近似平行四边形，且单元相对的两边凸向同一方向，最简单的强正规剖分是矩形分划)，有限元方程中各数值积分采用 $O(h^3)$ 二阶精度的高斯求积公式，那么在单元节点处有限元解的精度仍有 $O(h^2)$ 精度，或较好一点，但单元的形心处，$\partial u_h/\partial x$ 和 $\partial u_h/\partial y$ 却同时居有 $O(h^2)$ 的精度。同样也可利用邻点的高精度导数值进行线性内插和线性外推，以获得其它点处的导数佳值。

作为数值例子，考虑 $\Omega=[0,1]\times[0,1]$ 正方形区域上求解边值问题：

$$\begin{cases}\Delta u = -12x-2y+16x^2+54xy+16y^2-4x^3\\ \qquad -42x^2y-12xy^2-18y^3, \qquad (x,y)\in\Omega,\\ u=0, \qquad (x,y)\in\partial\Omega,\end{cases}$$

已知其准确解是 $u=x(1-x)y(1-y)(1+2x+7y)$。将Ω的每边分别作 $N=4$、6、8、10 等分，得到 4^2、6^2、8^2、10^2 个小正方形单元，用双线性元求解。用 1×1 Gauss 点求积公式，且用

$$E_S=\left\{\frac{1}{N^2}\sum_{z\in S}\left[\left(\frac{\partial(u-u_h)(z)}{\partial x}\right)^2+\left(\frac{\partial(u-u_h)(z)}{\partial y}\right)^2\right]\right\}^{1/2}$$

表示有限元导数在点集 S 上的平方平均误差；对于单元的中心的点集情形，用 E_G^1 表示；对于单元顶点集合情形，用 E_V^1 表示之；采用 2×2 Gauss 点求积公式且对应于单元形心点集的情形用E_G^2表示。它们的数值比较如下：

h	E_G^1	E_G^2	E_V^1
1/4	0.0550	0.0587	0.27
1/6	0.0250	0.0274	0.18
1/8	0.0140	0.0156	0.13
1/10	0.0091	0.0010	0.11

由这些数值可见，当 h 减小时，E_V^1 以速度 h 减小，而 E_G^1 以速度 h^2 减小；当 $h=1/4$ 时，$E_G^1=0.055$，这比 $h=1/10$ 时的 $E_V^1=0.11$还要小一倍。由此可见超收敛性的良好效果。另外，E_G^1和E_G^2的数值很接近，它表明在强正规分划时，采用 1×1Gauss 点求积公式即可保证有限元解具有足够的精度，没有多大必要采用更高精度的 Gauss 求积公式。

为了得到高精度的有限元位移值，仿本章§1的讨论，可利用网格 Ω_{2h} 和 Ω_h（将 Ω_{2h} 中每个单元的对边中点连接而成）节点上的有限元解 u_{2h} 和 u_h 作外推。

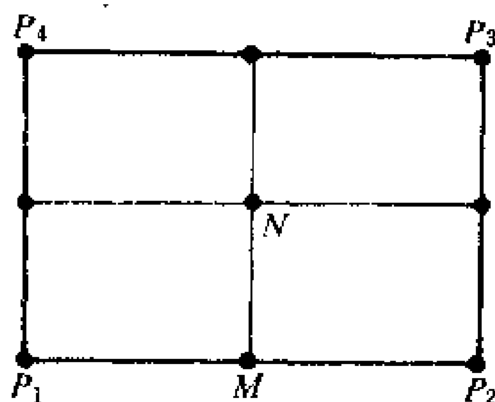


图 4.8

定理4.4[23] 设 $u\in H^3(\Omega)$，见图4.8，P_i 是 Ω_{2h} 与 Ω_h 分划的公共节点，M和N是 Ω_{2h} 分划中任一单元的边中点和单元中心，则

$$\begin{cases}\left|u(P_i)-\frac{1}{3}[4u_h(P_i)-u_{2h}(P_i)]\right|=O(h^3),\\ \left|u_h(M)+\frac{1}{6}[(u_h(P_1)-u_{2h}(P_1))+(u_h(P_2)\right.\\ \quad\left.-u_{2h}(P_2))]-u(M)\right|=O(h^3),\\ \left|u(N)-\sum_{i=1}^{4}\frac{1}{12}[4u_h(P_i)-u_{2h}(P_i)]\right|=O(h^3).\end{cases}\tag{4.15}$$

读者由定理 4.4 可知，适合于三角线性元情形的有限元解的渐近展式(4.1)对矩形双线性元情形也是适用的.

§3 高次元

为了提高有限元解的精度，可在单元上构造高次形函数. 显然，由于高次插值型函数对真解的逼近误差减小了，所以有限元解的精度也会提高.

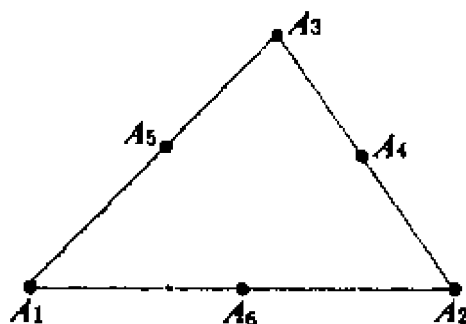


图 4.9

在一个三角形单元上构造二次插值型函数，须要有 6 个插值条件，为此须增加三个边中点，如图4.9所示. 若采用面积坐标，则这个二次函数是关于面积坐标的齐二次多项式，记为 $u_h(L_1, L_2, L_3)$，且符合插值型函数的特征. 若记三角单元对应节点处的基函数为 $N_i(L_1, L_2, L_3)$，则有

$$\begin{cases} N_i(A_j)=\delta_{ij}, \\ N_i(L_1,L_2,L_3)\text{是齐二次式}, \quad i,j=1,2,\cdots,6. \end{cases}$$

在直线 A_5A_6 上$\left(L_1=\frac{1}{2}\right)$, $N_1=0$, 这表明 N_1 含有$\left(L_1-\frac{1}{2}\right)$因子；在直线 $A_3A_4A_2$ 上($L_1=0$)，$N_1=0$，这表明 N_1 含有 L_1 因子；又因为 N_1 是齐二次式，所以必有形式 $N_1(L_1, L_2, L_3)=cL_1\left(L_1-\frac{1}{2}\right)$；再利用条件 $N_1(A_1)=1$, 得 $c=2$. 因此, $N_1=L_1(2L_1-1)$.

同样可得

$$N_2=L_2(2L_2-1), \qquad N_3=L_3(2L_3-1).$$

在直线 $A_1A_5A_3$ 上($L_2=0$)，$N_4=0$，这表明 N_4 含有 L_2 因子；在直线 $A_1A_6A_2$ 上($L_3=0$)，$N_4=0$，这表明 N_4 含有 L_3 因子；又因为 N_4 是关于面积坐标的齐二次式，所以有 $N_4=cL_2L_3$;

再利用条件 $N_4(A_4)=N_4\left(0,\frac{1}{2},\frac{1}{2}\right)=1$，有 $c=4$。因此，$N_4=4L_2L_3$。

同理有

$$N_5=4L_1L_3,\qquad N_6=4L_1L_2.$$

利用面积坐标和 x-y 坐标的关系式，即可得单元节点基函数 $N_i(x,y)(i=1,2,\cdots,6)$。对于单元顶点和边中点的基函数形状分别见图 4.10a)和图 4.10b)。

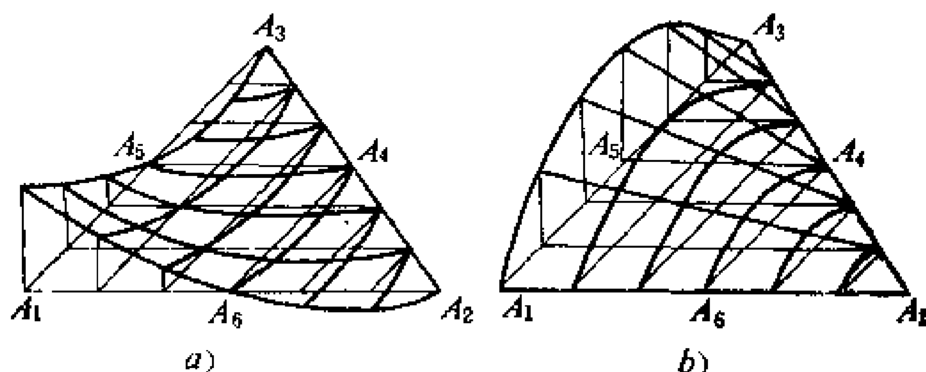


图 4.10

这样，任一单元 e 上的二次形函数为

$$u_{h,e}(x,y)=\sum_{i=1}^{6}u_iN_i(x,y),\quad (x,y)\in e.$$

再把单元节点基函数扩展为整体节点基函数 $\varphi_i(x,y)$，它在每个单元上是单元节点二次基函数；它在 i 节点的支集上非零，是整体连续的分片二次函数；它在 i 节点的支集外全为零.当然，所有整体节点基函数的线性组合为

$$u_h(x,y)=\sum_{i=1}^{N}u_i\varphi_i(x,y),$$

其中，u_i 是节点 i 处的函数近似值，N是单元顶节点和单元边中点的总数目，$u_h(x,y)$可作为解函数 $u(x,y)$ 的近似。显然，在整个 Ω_h 上，$u_h(x,y)$ 是分片二次的整体连续函数，$u_h(x,y)\in H^1(\Omega_h)$。

在标准正方形单元上构造双二次元也是很容易的。由(4.10)式可以清楚地看出，双线性单元节点基函数实际上是由一维线性单元节点基函数

$$\begin{cases} N_0(\xi)=1-\xi; \\ N_1(\xi)=\xi \end{cases} \text{和} \begin{cases} N_0(\eta)=1-\eta; \\ N_1(\eta)=\eta \end{cases}$$

组合乘积得到的，矩形双线性元有 4 个单元节点基函数。利用这一特点，将第一章§5的一维二次元基函数

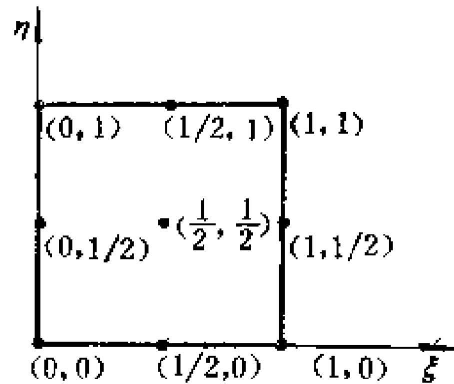


图 4.11

$$\begin{cases} N_0(\xi)=(2\xi-1)(\xi-1); \\ N_{\frac{1}{2}}(\xi)=4\xi(1-\xi); \\ N_1(\xi)=\xi(2\xi-1) \end{cases} \text{和} \begin{cases} N_0(\eta)=(2\eta-1)(\eta-1); \\ N_{\frac{1}{2}}(\eta)=4\eta(1-\eta); \\ N_1(\eta)=\eta(2\eta-1) \end{cases}$$

作组合乘积，就得到定义在标准正方形单元上的 9 个双二次元单元节点基函数(参见图 4.11)：

$$N_{0,0}(\xi,\eta)=(2\xi-1)(\xi-1)(2\eta-1)(\eta-1);$$

$$N_{0,1}(\xi,\eta)=(2\xi-1)(\xi-1)\eta(2\eta-1);$$

$$N_{1,0}(\xi,\eta)=\xi(2\xi-1)(2\eta-1)(\eta-1);$$

$$N_{1,1}(\xi,\eta)=\xi(2\xi-1)\eta(2\eta-1);$$

$$N_{\frac{1}{2},0}(\xi,\eta)=4\xi(1-\xi)(2\eta-1)(\eta-1);$$

$$N_{0,\frac{1}{2}}(\xi,\eta)=(2\xi-1)(\xi-1)4\eta(1-\eta);$$

$$N_{\frac{1}{2},\frac{1}{2}}(\xi,\eta)=16\xi(1-\xi)\eta(1-\eta);$$

$$N_{\frac{1}{2},1}(\xi,\eta)=4\xi(1-\xi)\eta(2\eta-1);$$

$$N_{1,\frac{1}{2}}(\xi,\eta)=4\xi(2\xi-1)\eta(1-\eta).$$

由此可见，构造双高次矩形元单元节点基函数有很强的规律性。

不论是三角分划还是四边形分划，人们都可以先在标准单元上构造出单元节点基函数；再利用仿射变换(2.13)式或(4.9)式

就能得到定义在每个实际单元上的单元节点基函数；把具有共同顶节点的所有单元上的单元节点基函数拼凑起来，就能得到定义在 Ω_h 上的整体节点基函数 $\varphi_i(x,y)$，这些 $\varphi_i(x,y)$ 是局部非零的，它在节点 i 的支集上非零，而在支集外恒为零；这样便可把所有整体节点基函数的线性组合

$$u_h(x,y)=\sum_{i=1}^{N}c_i\varphi_i(x,y) \tag{4.16}$$

作为 $u(x,y)$ 的近似函数；最后利用 Ritz-Galerkin 方法得到相应的有限元方程

$$\sum_{i=1}^{N}a(\varphi_i,\varphi_j)c_i=(f,\varphi_j),\quad j=1,2,\cdots,N. \tag{4.17}$$

由此可知，采用高次元的有限元方法与采用线性元的有限元方法具有统一的步骤。

如果在构造单元节点基函数时，只利用了插值节点处函数值的信息，那么这样构造出来的插值函数称作是 **Lagrange 型插值函数**. Lagrange 型 m 次元是定义在单元上的 m 次多项式插值函数。构造三角线性元要有 3 个插值条件，所以要 3 个插值节点；构造三角二次元要有 6 个插值条件，所以要有 6 个插值节点；构造三角三次元要有10个插值条件，所以要有10个插值节点；…。同样，矩形双线性元要有 4 个插值条件， 4 个插值节点；双二次元要有 9 个插值条件， 9 个插值节点；双三次元要有 16 个插值条件，16个插值节点；…。m 次 Lagrange 型插值函数虽然在每个单元上它的 m 阶导数不为零，但是就整体而言，特别是在单元公共边和节点处，它是连续函数且属于 $H^1(\Omega)$。在有限元方法中，若采用 Lagrange 型插值函数，则(4.16)式和(4.17)式中的 $c_i=u_i=u_h(x_i,y_i)$。

如果在构造单元节点基函数时，除了利用插值节点处的函数值信息，还利用了插值节点处的导数作为插值条件，那么这样构造出来的插值函数称作是 **Hermite 型插值函数**。Hermite 型 m 次元也是定义在单元上的 m 次多项式插值函数。例如，三角单元

上的三次 Hermite 型插值函数是一个关于 x 和 y 的三次多项式，需要10个插值条件，这可以由三个顶点上的函数值和两个偏导数值再加上三角单元形心处的函数值这10个条件来确定。又例如，矩形单元上的双三次 Hermite 型插值函数是一个关于 x 和 y 的双三次多项式，需要16个插值条件，这可以由4个顶点处给定的函数值、两个一阶偏导数的值和二阶混合导数的值，共16个条件来确定。由于使用了导数值作为插值条件，所以就整体而言，Hermite 型插值函数的光滑程度会有所提高。例如前面叙述的三次 Hermite 型插值函数是属于 $C^1(\Omega)$ 的。在有限元方法中，若采用 Hermite 型插值函数，则(4.16)式和(4.17)式中的 c_i 将是 $u_h(x, y)$ 在 i 节点处的函数值和导数值。为了不扰乱读者已清晰掌握的有限元方法的框架，为了节省篇幅，本篇不再详细介绍有关构造 Hermite 型插值函数的内容，有兴趣的读者可参阅[10]、[12]和[13]。

若在有限元方法中采用了 m 次元，则定义在凸区域 Ω 上的二阶椭圆边值问题的有限元解 u_h 适合如下误差估计[9]~[15]：

$$\|D^i(u-u_h)\|_0 \leqslant ch^{m+1-i}\|u\|_{m+1} \quad (i=0,1,\ u \in H^m(\Omega)),$$

$$\|D^i(u-u_h)\|_\infty \leqslant ch^{m+1-i}|\ln h|\,\|u\|_{m+1,\infty} \quad (i=0,1,\ u \in H^{m+1,\infty}(\Omega)).$$

显然，在一般情况下，采用高次元的有限元解的精度比采用线性元的有限元精度要好。但读者应该看到，采用高次元必须成倍地增设单元节点或插值条件，有限元方程组的阶数随之猛增，这样在形成和计算有限元方程组的过程中要花费的工作量也随之猛增；所以除特殊需要非用高次元不可之外，一般不必采用高次元，而采用线性元、网格细分、外推和利用超收敛性等措施，这样既可使计算工作量适度，又可提高有限元解的精度。当然，在弱解方程中含有较高阶导数时，线性元不能应用，此时必须采用高次元。

作为本章结束，我们必须指出，用有限元方法解椭圆边值问题时，产生误差或影响有限元解的精度的因素有很多种。例如，

离散化区域Ω_h对定解区域Ω逼近程度的好坏；单元剖分质量的好坏；网格细密程度；有限维函数u_h逼近真解的程度；有限元方程组中各个元素计算精度的好坏；总刚阵性质的好坏；解有限元方程组所选用的数值方法好坏；等等。只有全面地采取各种有效措施，才能非常有效地提高有限元解的精度。

第五章　一些有关的理论问题

本章主要介绍深入学习有限元方法必备的基础理论，内容仍以有限元方法展开，介绍变分法和Sobolev空间的有关知识，泛函分析和偏微分方程弱解适定性的基本问题，以及有限元误差估计等内容。本章内容可按学习需要穿插于第一章到第四章中学习，也可集中学习。

§1　变分法简介

变分问题是研究**泛函**的极值问题，这是一个很早就引起人们注意的问题。早在古埃及时期，就已提出“两固定点连线以直线为最短”的命题，阿基米德还将其作为直线的定义；古希腊人也提出并解决了“定周长线所围图形以圆的面积最大”的问题。那时，人们都是用古老的几何方法去研究变分问题的。

1696年，J. Bernaulli首先公开提出**最速降线问题**，并引起了一大批数学家的兴趣，从此数学进入了变分运算的年代。1744年，L. Euler综合了前人的工作，确立了**变分法**，从此变分运算及其研究迅速地发展起来。18世纪，由J. L. Lagrange和R. Gateaux进一步简化完善和发展了变分法，使其成为应用数学研究中的一个强有力的数学工具。

变分问题普遍地存在于自然现象中，因此在数学物理问题、力学、电学、弹塑性理论、数理经济学、最优控制论及系统理论等学科中的不少数学模型，在很大程度上都是用变分形式提出的。特别是有限元方法把变分形式作为数值解法的出发点，为此，我们先简要地介绍一些变分法的基本思想方法。

1.1 变分问题

最速降线问题 设A、B两点不在水平线上，在不考虑阻力的情况下，一小球沿哪一条光滑曲线从A点滑至B点所需时间最少？

不妨取A为原点，朝下置y轴(图5.1)。设曲线$\overset{\frown}{AB}$的方程为$y=y(x)$。因为仅受重力作用且初速为零，故滑行到点(x, y)处的速度$v^2=2gy$。于是小球滑行弧段ds所需的时间

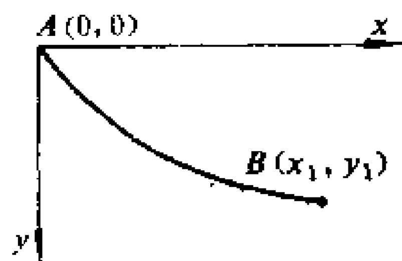


图5.1

$$dT=\frac{ds}{v}=\frac{\sqrt{1+\left(\frac{dy}{dx}\right)^2}}{\sqrt{2gy}}dx,$$

故总时间为

$$T=\int_0^{x_1}\sqrt{\frac{1+\left(\frac{dy}{dx}\right)^2}{2gy}}dx.$$

显然，T由曲线$y=y(x)$所决定，即是说，T是函数$y(x)$的函数。最速降线问题，就是要选择过A、B两定点的光滑曲线$y(x)$(不妨设$y(x)\in C^2[0,x_1]$)，当小球沿着这条光滑曲线滑行时，T取最小值。

最速降线问题的数学描述是：找一函数$y(x)\in M$，使有

$$T[y(x)]=\frac{1}{\sqrt{2g}}\int_0^{x_1}\sqrt{\frac{1+\left(\frac{dy}{dx}\right)^2}{y}}dx\Rightarrow\min. \qquad (5.1)$$

$$M:=\{y(x)\mid y(x)\in C^2[0,x_1], y(0)=0, y(x_1)=y_1\}.$$

最小面积问题 在一固定框架上张一曲面，问哪一个曲面的面积最小？

设固定框架在 xoy 平面上的投影为 Ω，Ω 的边界记为 $\partial\Omega$，曲面位移记为 $z(x, y)$，$z(x, y)$在 $\partial\Omega$ 上的值为$\varphi(x, y)$，见图 5.2. 易知曲面面积 S的计算公式是

$$S[z(x,y)]=\int_\Omega\left[1+\left(\frac{\partial z}{\partial x}\right)^2+\left(\frac{\partial z}{\partial y}\right)^2\right]^{1/2}dxdy$$

$$=\int_\Omega(1+|\nabla z|^2)^{1/2}dxdy.$$

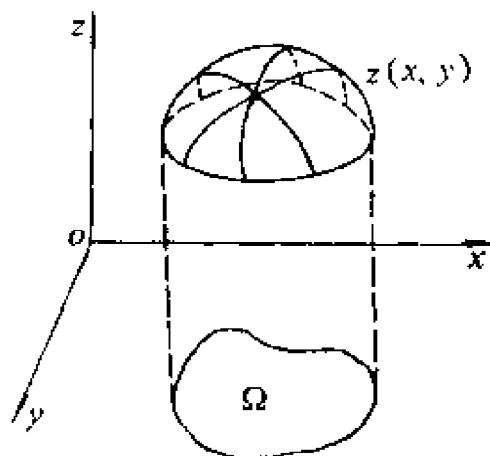


图5.2

显然面积 S 是函数$z(x,y)$的函数，不妨设 $z(x,y)\in C^2(\Omega)$，最小面积问题就是要在那些属于 $C^2(\Omega)$且在 $\partial\Omega$ 上函数值为 $\varphi(x,y)$的光滑曲面中，选出一个曲面$z(x,y)$使其面积最小。

最小面积问题的数学描述是：找一函数$z(x,y)\in M$，使有

$$S[z(x,y)]=\int_\Omega(1+|\nabla z|^2)^{1/2}dxdy\Rightarrow\min. \qquad (5.2)$$

$$M:=\{z(x,y)\,|\,z(x,y)\in C^2(\Omega),\ z(x,y)|_{\partial\Omega}=\varphi(x,y)\}.$$

上面两个典型问题中，M是**可供选择的函数类**或称为**容许函数类**，其中的函数据实际问题的要求须有光滑程度及边值的限定；$T[y]$和$S[z]$都是函数的函数，对M中的每一确定的函数，它都与一确定的实数对应，这些函数的函数称为**泛函**，泛函的定义域是函数类，值域是实数域R；求泛函的极值问题称为**变分问题**。

1.2 变分法

设 $u_0(x)\in M$，$u(x)$是容许函数类M中的任一函数，如果 $J(u_0)\leqslant J(u)$（或$J(u_0)\geqslant J(u)$），则称泛函$J(u)$在 u_0 处取得了**整体极小值**（或**整体极大值**）；如果作为比较的函数$u(x)$仅限于函

数$u_0(x)$的某一邻域中，则称$J(u_0)$在u_0处取得了**局部极小值**(或**局部极大值**)。泛函的整体(或局部)极值类似于普通函数的整体(或局部)极值。就变分问题来说，可据实际问题的需要，人们要寻求泛函的极值或驻值。

现回忆一下求函数极值的办法。设函数$f(x)$在x_0处取得极值$f(x_0)$，自变量取得增量Δx后，函数$f(x)$在x_0处取得极值$f(x_0)$应满足条件

$$\lim_{\Delta x\to 0}\frac{f(x_0+\Delta x)-f(x_0)}{\Delta x}=f'(x_0)=0,\text{ 和}f''(x_0)>0.$$

另一种办法是，极值点x_0处取得的增量Δx可表示为$\Delta x=\lambda v(\lambda\in R,\forall v\in R)$，此时函数$f(x)$可改写为$f(x_0+\lambda v)$，显然它是一个关于实参数$\lambda$的函数且$\lambda=0$时才能取得极值$f(x_0)$，根据微分法的基本知识易知，$f(x)$在$x_0$处取得极值应满足条件

$$f'_\lambda(x_0+\lambda v)|_{\lambda=0}=0,\text{ 和}f''_\lambda(x_0+\lambda v)|_{\lambda=0}>0,\quad\forall v\in R.$$

同样，泛函求极值也可采用上述两种办法，不过第二种办法更简单易行。为此设泛函$J(u)$在$u_0(x)$处取得极值，容许函数类M中的任一函数$u(x)$可看作是$u_0(x)$经一扰动函数$v(x)$作用的结果，即有

$$u(x)=u_0(x)+\lambda v(x)\in M.$$

注意，此时u_0、u都属于容许函数类，故扰动函数$v(x)$的基本特征是：在固定边界处恒有$v(x)\equiv 0$。记$v(x)\in M_0$，M_0称为**试探函数空间**。于是对容许函数类M中任一函数$u(x)$所确定的泛函$J(u)$可写成下列形式

$$J(u)=J(u_0+\lambda v)=\varphi(\lambda).$$

读者应特别注意，$J(u_0+\lambda v)$是一个关于实参数λ的实值函数且在$\lambda=0$时才取得极值$J(u_0)$，故运用经典微分法知识，泛函$J(u)$在u_0处取得极值应满足条件

$$J'_\lambda(u_0+\lambda v)|_{\lambda=0}=0, \tag{5.3}$$

和

$$J''_{\lambda}(u_0+\lambda v)|_{\lambda=0}>0. \tag{5.4}$$

条件(5.3)又称为 **Euler 必要条件**，满足 Euler 必要条件的解可能是驻点而不是极值点，只有同时满足条件(5.3)和(5.4)的解才是极值解。

顺便指出，利用条件(5.3)可得到一个积分方程，再利用变分预备定理，又可得到一个微分方程，这个由Euler必要条件获得的微分方程称为 **Euler 方程**。下面先介绍将经常用到的变分预备定理。

变分预备定理 设$u(x)\in C^0[a,b]$且记

$C_0^1[a,b]:=\{v(x)|v(x)\in C^1[a,b],\ v(a)=v(b)=0\}$，

若

$\int_a^b u(x)v(x)dx=0,\ \forall v(x)\in C_0^1[a,b]$，

则
$$u(x)\equiv 0,\ \forall x\in[a,b].$$

证明 用反证法。设$\bar{x}\in[a,b]$使$u(\bar{x})\neq 0$，不妨设$u(\bar{x})>0$，由于$u(x)$的连续性，则必存在区间$[\xi_1,\xi_2]\subset[a,b]$且$\bar{x}\in[\xi_1,\xi_2]$，使下式成立：

$$u(x)>0,\ x\in[\xi_1,\xi_2].$$

现构造

$$v(x)=\begin{cases}(x-\xi_1)^2(x-\xi_2)^2, & x\in(\xi_1,\xi_2),\\ 0, & \text{别处},\end{cases}$$

显然$v(x)\in C_0^1[a,b]$，但积分

$$\int_a^b u(x)v(x)\ dx=\int_{\xi_1}^{\xi_2}u(x)v(x)dx>0,$$

所以与假设矛盾，这就证明了$u(x)\equiv 0$。 □

上述变分预备定理还可推广为二维情形：

设有界闭区域$\bar{\Omega}$，$u(x,y)\in C^0(\bar{\Omega})$，记

$C_0^1(\bar{\Omega}):=\{v(x,y)|v(x,y)\in C^1(\bar{\Omega}),v|_{\partial\Omega}=0\}$，

若

$$\int_{\Omega} u(x,y)v(x,y)dxdy=0,\ \forall v(x,y)\in C_0^1(\bar{\Omega}),$$

则 $u(x,y)\equiv 0,\ \forall (x,y)\in\bar{\Omega}$。

1.3 Euler 方程

考虑用变分法求解最速降线变分问题(5.1)，据题义易知，容许函数类和试探函数空间分别是

$$M:=\{y(x)\mid y(x)\in C^2[0,x_1],\ y(0)=0,\ y(x_1)=y_1\},$$

$$M_0:=\{v(x)\mid v(x)\in C^2[0,x_1],\ v(0)=v(x_1)=0\};$$

设变分问题(5.1)在$y(x)$处取得极值，经扰动函数作用后，$y(x)$和$y'(x)$应分别变为$y(x)+\lambda v(x)$和$y'(x)+\lambda v'(x)$，从而(5.1)式表示为

$$\begin{cases} T(y+\lambda v)=\dfrac{1}{\sqrt{2g}}\displaystyle\int_0^{x_1}\sqrt{\dfrac{1+(y'+\lambda v')^2}{(y+\lambda v)}}dx\Rightarrow\min, \\ y+\lambda v\in M,\ \forall v\in M_0. \end{cases}$$

利用 Euler 必要条件(5.3)和分部积分法，有

$$T'_\lambda(y+\lambda v)\Big|_{\lambda=0}=\frac{1}{\sqrt{2g}}\int_0^{x_1}\left[\frac{y'}{\sqrt{y(1+y'^2)}}v'-\frac{1}{2y}\sqrt{\frac{1+y'^2}{y}}v\right]dx$$

$$=\frac{-1}{\sqrt{2g}}\int_0^{x_1}\left\{\frac{1}{2y}\sqrt{\frac{1+y'^2}{y}}+\frac{d}{dx}\left[\frac{y'}{\sqrt{y(1+y'^2)}}\right]\right\}v\,dx=0, \tag{5.5}$$

再利用变分预备定理，得

$$\frac{1}{2y}\sqrt{\frac{1+y'^2}{y}}+\frac{d}{dx}\left[\frac{y'}{\sqrt{y(1+y'^2)}}\right]=0.$$

这就是最速降线变分问题的解所适合的 Euler 方程，它可改写为

$$\frac{d}{dx}\left[\frac{y'^2}{\sqrt{y(1+y'^2)}}-\sqrt{\frac{1+y'^2}{y}}\right]=\frac{d}{dx}\left[\frac{-1}{\sqrt{y(1+y'^2)}}\right]=0,$$

积分一次，有 $y(1+y'^2)=c_1$，并设$y'=\text{ctg}\,t$ 则得

$$y=\frac{c_1}{2}(1-\cos 2t),$$

$$x=\frac{c_1}{2}(2t-\sin 2t)+c_2.$$

最后利用初始条件$x=0$，$y=0$，并设$\theta=2t$，就得最速降线变分问题的解

$$x=\frac{c}{2}(\theta-\sin\theta),$$

$$y=\frac{c}{2}(1-\cos\theta)$$

是一圆滚线（摆线）。

最速降线（摆线）在实际中有重要应用，将其作为齿轮轮廓线，齿轮就耐磨；钟表业中也使用摆线，当摆锤按摆线摆动时，钟摆周期会更准确。

一般地，求含有一阶导数的变分问题

$$\begin{cases} J(y)=\int_a^b F(x,y,y')\,dx\Rightarrow\min, \\ \forall y\in M:\ =\{y(x)\,|\,y(x)\in C^2[a,b],y(a)=y_1,y(b)=y_2\} \end{cases} \tag{5.6}$$

所满足的Euler方程也很容易，此时试探函数空间为

$$M_0:\ =\{v(x)\,|\,v(x)\in C^2[a,b],v(a)=v(b)=0\}.$$

若记

$$\bar{F}=F(x,y+\lambda v,y'+\lambda v')=F(x,\bar{y},\bar{y}'),$$

$$\bar{y}=y+\lambda v,\ \bar{y}'=y'+\lambda v',$$

那么，利用Euler必要条件并使用分部积分法，就有

$$J'(\lambda)\Big|_{\lambda=0}=\frac{dJ(y+\lambda v)}{d\lambda}\Big|_{\lambda=0}$$

$$=\int_a^b\left[\frac{\partial\bar{F}}{\partial\bar{y}}\frac{d\bar{y}}{d\lambda}+\frac{\partial\bar{F}}{\partial\bar{y}'}\frac{d\bar{y}'}{d\lambda}\right]_{\lambda=0}dx$$

$$=\int_a^b\left[\frac{\partial F}{\partial y}v+\frac{\partial F}{\partial y'}v'\right]dx$$

$$= \int_a^b \left[\frac{\partial F}{\partial y} - \frac{d}{dx}\left(\frac{\partial F}{\partial y'}\right)\right] v\,dx = 0. \qquad (5.7)$$

最后，利用变分预备定理，得到变分问题(5.6)所满足的Euler方程：

$$\frac{\partial F}{\partial y} - \frac{d}{dx}\left(\frac{\partial F}{\partial y'}\right) = 0,$$

$$y(a) = y_1, \quad y(b) = y_2.$$

对于含有高阶导数的泛函

$$J(y) = \int_a^b F(x, y, y', \cdots, y^{(m)})\,dx,$$

必须给定容许函数$y(x)$的边界条件：

$$y(a) = \alpha_0, y'(a) = \alpha_1, \cdots, y^{(m-1)}(a) = \alpha_{m-1},$$
$$y(b) = \beta_0, y'(b) = \beta_1, \cdots, y^{(m-1)}(b) = \beta_{m-1}.$$

由实现极值的Euler必要条件，读者不难推出其相应的Euler方程是

$$\begin{cases} F_y - \dfrac{d}{dx} F_{y'} + \dfrac{d^2}{dx^2} F_{y''} - \cdots + (-1)^m \dfrac{d^m}{dx^m} F_{y^{(m)}} = 0, \\ y^{(i)}(a) = \alpha_i, \quad y^{(i)}(b) = \beta_i, \quad i = 0, 1, \cdots, m-1. \end{cases}$$

用变分方法处理二维变分问题也很方便。

例如求变分问题

$$\begin{cases} J(u) = \displaystyle\int_\Omega F(x, y, u, u_x, u_y)\,dx\,dy \Rightarrow \min, \\ u(x,y) \in M: = \{u(x,y) \in C^2(\Omega) \mid u|_{\partial\Omega} = g(x,y)\} \end{cases} \qquad (5.8)$$

所满足的Euler方程，此时试验函数空间为

$$M_0: = \{v(x,y) \in C^2(\Omega) \mid v|_{\partial\Omega} = 0\}.$$

若泛函$J(u)$在$u(x,y)$处取得极值，那么u、u_x和u_y经扰动函数作用后相应地变为$u+\lambda v$、$u_x+\lambda v_x$和$u_y+\lambda u_y$，从而M中任一容许函数的变分问题(5.8)可改写为

$$\begin{cases} J(u+\lambda v) = \int_\Omega F(x,y,u+\lambda v, u_x+\lambda v_x, u_y+\lambda v_y)dxdy \Rightarrow \min, \\ u+\lambda v \in M, \quad \forall v \in M_0. \end{cases}$$

利用 Green 公式

$$\begin{cases} \int_\Omega uv_x dxdy + \int_\Omega u_x v dxdy = \int_{\partial\Omega} uvn_x ds, \\ \int_\Omega uv_y dxdy + \int_\Omega u_y v dxdy = \int_{\partial\Omega} uvn_y ds, \end{cases}$$

(其中 n_x 和 n_y 是单位法向量的x和y分量)，和 Euler 必要条件，有

$$J'(u+\lambda v)\Big|_{\lambda=0} = \int_\Omega \left[\frac{\partial F}{\partial u}v + \frac{\partial F}{\partial u_x}v_x + \frac{\partial F}{\partial u_y}v_y\right]dxdy$$

$$= \int_\Omega \left[\frac{\partial F}{\partial u} - \frac{\partial}{\partial x}\left(\frac{\partial F}{\partial u_x}\right) - \frac{\partial}{\partial y}\left(\frac{\partial F}{\partial u_y}\right)\right]vdxdy$$

$$+ \int_{\partial\Omega}\left(\frac{\partial F}{\partial u_x}n_x + \frac{\partial F}{\partial u_y}n_y\right)vds = 0;$$

又因 $v \in M_0$，即 $v|_{\partial\Omega} = 0$，所以有

$$\int_\Omega \left[\frac{\partial F}{\partial u} - \frac{\partial}{\partial x}\left(\frac{\partial F}{\partial u_x}\right) - \frac{\partial}{\partial y}\left(\frac{\partial F}{\partial u_y}\right)\right]vdxdy = 0. \qquad (5.9)$$

最后利用变分预备定理，并注意到 $u|_{\partial\Omega} = g(x,y)$，就得到变分问题(5.8)所满足的Euler 方程

$$\begin{cases} \frac{\partial F}{\partial u} - \frac{\partial}{\partial x}\left(\frac{\partial F}{\partial u_x}\right) - \frac{\partial}{\partial y}\left(\frac{\partial F}{\partial u_y}\right) = 0, \quad (x,y) \in \Omega, \\ u|_{\partial\Omega} = g(x,y). \end{cases}$$

作为例子，变分问题

$$\begin{cases} J(u) = \int_\Omega \left[\left(\frac{\partial u}{\partial x}\right)^2 + \left(\frac{\partial u}{\partial y}\right)^2 - 2uf\right]dxdy \Rightarrow \min, \\ M: = \{u \in C^2(\Omega) \mid u|_{\partial\Omega} = g(x,y)\} \end{cases}$$

的解所适合的 Euler 方程是 Poisson 方程

$$\begin{cases} -\Delta u = f(x,y), & (x,y)\in\Omega, \\ u|_{\partial\Omega} = g(x,y). \end{cases}$$

同样，也很容易求得变分问题(5.2)所适合的Euler方程是

$$\begin{cases} \sum_{i=1}^{2} -\dfrac{\partial}{\partial x_i}\left[\dfrac{\partial z/\partial x_i}{(1+|\nabla z|^2)^{1/2}}\right] = 0, & (x,y)\in\Omega, \\ z|_{\partial\Omega} = \varphi(x,y). \end{cases}$$

再看一个具有多种边值的例子，求变分问题

$$\begin{cases} J(u) = \int_\Omega F(u,u_x,u_y)\,dx\,dy + \int_{\Gamma_2} G(u)\,ds \Rightarrow \min, \\ M: = \{u(x,y)\in C^2(\Omega)\,|\,u|_{\Gamma_1} = g(x,y)\} \end{cases} \tag{5.10}$$

所适合的 Euler 方程，其中 $\partial\Omega = \Gamma_1 \cup \Gamma_2$ 如图 5.3 所示。



图5.3

据扰动函数的特点：在固定边界处，扰动函数值必须为零，在其它边界部分，其值任意；易知试探函数空间

$$M_0: = \{v(x,y)\in C^2(\Omega)\,|\,v(x,y)|_{\Gamma_1} = 0\}.$$

完全类似于前面的推导，泛函可表示为

$$J(u+\lambda v) = \int_\Omega F(u+\lambda v, u_x+\lambda v_x, u_y+\lambda v_y)\,dx\,dy + \int_{\Gamma_2} G(u+\lambda v)\,ds,$$

应用 Euler 必要条件和 Green 公式，则有

$$\begin{aligned} J'(u+\lambda v)\Big|_{\lambda=0} &= \int_\Omega\left[\frac{\partial F}{\partial u}v + \frac{\partial F}{\partial u_x}v_x + \frac{\partial F}{\partial u_y}v_y\right]dx\,dy + \int_{\Gamma_2}\frac{dG}{du}v\,ds \\ &= \int_\Omega\left[\frac{\partial F}{\partial u} - \frac{\partial}{\partial x}\left(\frac{\partial F}{\partial u_x}\right) - \frac{\partial}{\partial y}\left(\frac{\partial F}{\partial u_y}\right)\right]v\,dx\,dy + \int_{\Gamma_2}\frac{dG}{du}v\,ds \\ &\quad + \int_{\partial\Omega}\left[\frac{\partial F}{\partial u_x}n_x + \frac{\partial F}{\partial u_y}n_y\right]v\,ds = 0, \end{aligned}$$

注意 $v|_{\Gamma_1}=0$，上式则改写为

$$\int_{\Omega}\left[\frac{\partial F}{\partial u}-\frac{\partial}{\partial x}\left(\frac{\partial F}{\partial u_x}\right)-\frac{\partial}{\partial y}\left(\frac{\partial F}{\partial u_y}\right)\right]v\,dx\,dy$$

$$+\int_{\Gamma_2}\left[\frac{dG}{du}+\frac{\partial F}{\partial u_x}n_x+\frac{\partial F}{\partial u_y}n_y\right]v\,ds=0. \tag{5.11}$$

注意对 $\forall v\in M_0$，(5.11)式都是成立的。M_0 中的试探函数 v 可分为二类：第一类，$v|_{\Gamma_1}=v|_{\Gamma_2}=0$，第二类，$v|_{\Gamma_1}=0, v|_{\Gamma_2}\neq 0$。不妨先取第一类试探函数，这样由(5.11)式得到

$$\frac{\partial F}{\partial u}-\frac{\partial}{\partial x}\left(\frac{\partial F}{\partial u_x}\right)-\frac{\partial}{\partial y}\left(\frac{\partial F}{\partial u_y}\right)=0,\ (x,y)\in\Omega$$

在 Ω 上几乎处处成立。所以问题转化为

$$\int_{\Gamma_2}\left[\frac{dG}{du}+\frac{\partial F}{\partial u_x}n_x+\frac{\partial F}{\partial u_y}n_y\right]v\,ds=0,\ \forall v\in M_0.$$

再取第二类试探函数就得到

$$\frac{dG}{du}+\frac{\partial F}{\partial u_x}n_x+\frac{\partial F}{\partial u_y}n_y=0,\ (x,y)\in\Gamma_2$$

在 Γ_2 上几乎处处成立。联同 $u\in M$ 必须满足的条件 $u|_{\Gamma_1}=g$，就得到变分问题(5.10)的解所满足的 Euler 方程

$$\begin{cases}\dfrac{\partial F}{\partial u}-\dfrac{\partial}{\partial x}\left(\dfrac{\partial F}{\partial u_x}\right)-\dfrac{\partial}{\partial y}\left(\dfrac{\partial F}{\partial u_y}\right)=0,\ (x,y)\in\Omega,\\[2mm] \dfrac{dG}{du}+\dfrac{\partial F}{\partial u_x}n_x+\dfrac{\partial F}{\partial u_y}n_y=0,\ (x,y)\in\Gamma_2,\\[2mm] u=g,\ (x,y)\in\Gamma_1.\end{cases}$$

1.4 几点说明

1. 本节中列出的几种变分问题都可以利用**极小位能原理**获得。极小位能原理表明，一切处于静态平衡的物体和系统都具有极小位能。所以极小位能原理是能量守恒律的表现，利用这一能量原理，读者可以容易地建立各学科中存在的变分问题（参见第一章§1）。

2. 求解变分问题使用变分法，那么 Euler 必要条件一定可转化为

$$\int_{\Omega}(Lu-f)v\,dx\,dy+\int_{\Gamma_2}(Gu-\beta)v\,ds=0 \qquad (5.12)$$

的形式，式中 L 是定义在 Ω 上的**微分算子**，G 是定义在第二、三类边值边界 Γ_2 上的**边值微分算子**，v 是试探函数（参见(5.5)～(5.11)式）；再利用变分预备定理就能把能量积分形式(5.12)转化为微分方程（Euler 方程）；因此可以断言：椭圆型微分方程一定可以转化为能量积分形式，且这种形式转换的关键就是(5.12)式。

3. 能量积分形式和微分方程形式是描述事物性态的两种不同的数学形式；能量积分形式要求解的光滑程度低，例如微分方程中要求解 $u\in C^{2m}$，而能量积分形式中只要求 m 阶偏导数可积分；能量积分形式中还包含了第二、三类边值条件，这种数学形式对于处理复杂的边值条件是有益的。特别地，能量积分形式作为有限元方法解微分方程的出发点，从而使有限元方法保持了上述种种优势。

4. 为了学习有限元方法，本节只是简略地介绍了一些有关泛函、变分问题、变分法的基本概念及其相互关系，有关变分法及其应用的深入研究，有兴趣的读者还可参阅[2～4]。

§2 Sobolev空间简介

2.1 引进和扩充函数类的必要性

在本章§1和第一章§1中，我们讨论了当 $u\in C^2$ 时，变分问题、变分方程和微分方程边值问题这三种数学形式的相互转换关系。例如弦平衡问题，其**变分问题**是

$$J(u)=\int_0^l\left[\frac{1}{2}T\cdot(u')^2-uf\right]dx\Rightarrow\min \quad u\in M;$$

利用变分法，得变分方程

$$\int_0^l Tu'v'dx = \int_0^l fv dx, \quad \forall v \in M_0;$$

再利用分部积分公式，得**虚功原理**表现形式：

$$\int_0^l [-Tu'' - f] v dx = 0, \quad \forall v \in M_0;$$

最后利用变分预备定理得**Euler 方程**

$$\begin{cases} -Tu'' = f, \\ u(0) = u(l) = 0. \end{cases}$$

在实际问题中，外力 f 可能是间断力甚至是集中力，此时的变分问题和变分方程仍然是有意义的，但微分方程的边值问题却没有属于$C^2[0,l]$的经典解，因此限制$u \in C^2[0,l]$来考虑弦平衡问题是不尽合理的。另一方面，只需要条件 $u \in L_2$ 和 $u' \in L_2$，变分问题和变分方程就有意义，这就要求人们进一步扩大函数类且在这个更广泛的函数类中研究变分问题、变分方程和微分方程边值问题。

顺便指出，函数类 $L_2[a,b]$是指$|u(x)|^2$在$[a,b]$上可积分的那些函数 u 的全体，即

$$L_2[a,b]: = \{u(x) \mid \int_a^b |u(x)|^2 dx < +\infty\};$$

u 的模记为$\|u\|_{L_2}$或$\|u\|_0$，即

$$\|u\|_{L_2} = \|u\|_0 = \left[\int_a^b |u|^2 dx\right]^{1/2}.$$

$\|u\|_0$可反映某函数所具有的某种能量（例如势能）；$\|u_1 - u\|_0$可反映 u_1和（$-u_2$）叠加后所具有的能量，从数学上看$\|u_1 - u_2\|_0$则表示了这两个函数之间的距离。显然，连续可微函数，分段连续函数，阶梯函数都属于$L_2[a,b]$这个函数类。

对于u、$v \in L_2[a,b]$，定义它们的内积

$$(u,v): = \int_a^b u(x)v(x)dx,$$

可以证明著名的 Schwartz **不等式**：

$$|(u,v)|\leqslant\|u\|_0\|v\|_0 \tag{5.13}$$

和三角不等式：

$$\|u+v\|_0\leqslant\|u\|_0+\|v\|_0 \tag{5.14}$$

成立。事实上，对任何实数 λ，恒有

$$\int_a^b(\lambda u+v)^2dx=\lambda^2\int_a^b u^2dx+2\lambda\int_a^b uv\,dx+\int_a^b v^2dx\geqslant 0,$$

从而对上面的关于 λ 的二次三项式中，恒有其判别式非正，即有

$$\left(\int_a^b uv\,dx\right)^2-\int_a^b u^2dx\cdot\int_a^b v^2dx\leqslant 0,$$

从而证得(5.13)式。其次，利用(5.13)式有

$$\|u+v\|_0^2=\int_a^b(u^2+2uv+v^2)dx=\|u\|_0^2+2(u,v)+\|v\|_0^2$$

$$\leqslant\|u\|_0^2+2\|u\|_0\|v\|_0+\|v\|_0^2=(\|u\|_0+\|v\|_0)^2,$$

两边开方，就推出(5.14)式。

对级数求和利用 Schwartz 不等式，可得到 Cauchy 不等式

$$\left|\sum_{i=1}^n a_ib_i\right|\leqslant\left(\sum_{i=1}^n a_i^2\right)^{1/2}\left(\sum_{i=1}^n b_i^2\right)^{1/2}.$$

另外，变分预备定理也有更弱的形式：

设 $u(x)\in L_2[a,b]$，记

$$C_0^\infty[a,b]:=\left\{v(x)\,\middle|\,v(x)\in C^\infty[a,b],\frac{d^iv}{dx}(a)=\frac{d^iv}{dx}(b)=0,\ i=0,1,2,\cdots\right\}.$$

若 $\int_a^b u(x)v(x)dx=0,\quad\forall v(x)\in C_0^\infty[a,b]$，

则 $u(x)$ 在 $[a,b]$ 上几乎处处为零。

2.2 广义导数的定义和性质

为了在更广泛的函数类中研究变分问题（或积分方程）与微分问题（或微分方程）之间的转换，最关键的是它们满足分部积

分公式，且需推广导数的概念。

设$\varphi\in C_0^\infty[a,b]$，那么当$f\in C^k[a,b]$时，分部积分公式成立，即

$$\int_a^b f'(x)\varphi(x)\,dx=-\int_a^b f(x)\varphi'(x)\,dx,\quad \forall\varphi\in C_0^\infty[a,b],$$

$$\int_a^b f^{(k)}(x)\varphi(x)\,dx=(-1)^k\int_a^b f(x)\varphi^{(k)}(x)\,dx,\quad \forall\varphi\in C_0^\infty[a,b].$$

不难理解，只要在更弱的条件下使分部积分公式成立，就可推广导数的概念。

定义1 设$f(x)\in L_2[a,b]$，若存在$g(x)\in L_2[a,b]$,使等式

$$\int_a^b g(x)\varphi(x)\,dx=-\int_a^b f(x)\varphi'(x)\,dx,\quad \forall\varphi(x)\in C_0^\infty[a,b]$$

恒成立，则称$g(x)$是$f(x)$的**一阶广义导数**，记为$f'(x)$或Df。

广义导数有下面几个性质。

性质一 经典导数必是广义导数，但反之不真。

事实上，若设$f'(x)$是$f(x)$的经典导数，则必满足分部积分公式，所以$f'(x)\in L_2[a,b]$也是$f(x)$的广义导数。若设$f'(x)$是$f(x)$的广义导数，因$f(x)\in L_2\not\subset C^1$，故$f'(x)$不是$f(x)$的经典导数。

例如，对于$f(x)=|x|,x\in[-1,1]$，则$f(x)$在$[-1,1]$上广义可导，但不存在经典导数。

对于这个例子，人们可直接利用广义导数的概念求出$f(x)=|x|$的广义导数，此时有

$$-\int_{-1}^1 f(x)\varphi'(x)\,dx=-\int_{-1}^1 |x|\varphi'(x)\,dx$$

$$=\int_{-1}^0 x\varphi'(x)\,dx-\int_0^1 x\varphi'(x)\,dx=-\int_{-1}^0\varphi(x)\,dx+\int_0^1\varphi(x)\,dx$$

$$=\int_{-1}^1 g(x)\varphi(x)\,dx,\qquad \forall\varphi(x)\in C_0^\infty[-1,1],$$

式中，$g(x)=\begin{cases}1, & 0\leqslant x\leqslant 1\\ -1, & -1\leqslant x<0\end{cases}$或$g(x)=\begin{cases}1, & 0<x\leqslant 1\\ -1, & -1\leqslant x\leqslant 0\end{cases}$

就是$f(x)$的广义导数。

由这个例子，还容易看出广义导数的另一个性质。

性质二 同一函数$f(x)$的广义导数是在几乎处处的意义下唯一的。

事实上，若设$f_1'(x)$和$f_2'(x)$都是$f(x)$的广义导数，那么由广义导数的定义，有

$$\int_a^b f(x)\varphi'(x)\,dx = -\int_a^b f_1'(x)\varphi(x)\,dx,\quad \forall \varphi(x)\in C_0^\infty[a,b],$$

$$\int_a^b f(x)\varphi'(x)\,dx = -\int_a^b f_2'(x)\varphi(x)\,dx,\quad \forall \varphi(x)\in C_0^\infty[a,b],$$

从而

$$\int_a^b (f_1'(x)-f_2'(x))\varphi(x)\,dx = 0,\quad \forall \varphi(x)\in C_0^\infty[a,b],$$

再应用变分预备定理就证得性质二正确。

性质三 若$f\in L_2[a,b]$，则广义导数$f'(x)$不一定存在。

事实上，阶梯函数

$$f(x)=\begin{cases}0, & -1\leqslant x\leqslant 0\\ 1, & 0<x\leqslant 1\end{cases}$$

的广义导数不存在。

因为，若设$f(x)$有广义导数$g(x)$，则有

$$\int_{-1}^1 g(x)\varphi(x)\,dx = -\int_{-1}^1 f(x)\varphi'(x)\,dx$$

$$= -\int_0^1 \varphi'(x)\,dx = -\varphi(x)\big|_0^1 = \varphi(0),\quad \forall \varphi(x)\in C_0^\infty[-1,1]。$$

特别地，取

$$\varphi(x)=\varphi_\varepsilon(x)=\begin{cases}\exp\left(-\dfrac{1}{1-(x/\varepsilon)^2}\right), & \text{当}\ |x|\leqslant\varepsilon\ (0<\varepsilon<1)\text{时},\\ 0, & \text{在别处},\end{cases}$$

那么$\varphi_\varepsilon(x)\in C_0^\infty[-1,1]$。一方面

$$\left|\int_{-1}^1 g(x)\varphi(x)\,dx\right| = |\varphi_\varepsilon(0)| = e^{-1}>0;$$

另一方面

$$\left|\int_{-1}^{1} g(x)\varphi(x)\,dx\right| \leqslant \|g(x)\|_0\|\varphi(x)\|_0$$

$$= \|g(x)\|_0\left(\int_{-\varepsilon}^{\varepsilon}\left|\exp\left(\frac{-1}{1-(x/\varepsilon)^2}\right)\right|^2 dx\right)^{1/2}$$

$$= \|g(x)\|_0\left(2\varepsilon\int_0^1\left|\exp\left(\frac{-1}{1-t^2}\right)\right|^2 dt\right)^{1/2} \to 0,\quad \varepsilon\to 0,$$

导致矛盾。故知$g(x)\notin L_2[-1,1]$，所以阶梯函数没有广义导数。

仿照一阶广义导数的定义，可定义高阶广义导数。

定义2 设$f(x)\in L_2[a,b]$，若存在$g(x)\in L_2[a,b]$，使有

$$\int_a^b f(x)\varphi^{(k)}(x)\,dx = (-1)^k\int_a^b g(x)\varphi(x)\,dx,\quad \forall\varphi(x)\in C_0^\infty[a,b]$$

成立，则称$g(x)$是$f(x)$的 **k 阶广义导数**，记为$f^{(k)}(x)$或$D^k f$。

2.3 Sobolev 空间$H^k(a,b)$和$H_0^k(a,b)$

那些具有直到k阶广义导数的函数的全体，记为

$$H^k(a,b):=\{f(x)\,|\,f^{(\alpha)}(x)\in L_2(a,b),\ \alpha=0,1,\cdots,k\}.$$

显然$H^k(a,b)$是一线性空间。

若再定义**内积**和**模**分别为

$$(u,v)_k:=\sum_{\alpha=0}^{k}\int_a^b f^{(\alpha)}g^{(\alpha)}dx,$$

$$\|f\|_k=\|f\|_{k,2,[a,b]}:=\left(\sum_{\alpha=0}^{k}\int_a^b (f^{(\alpha)})^2 dx\right)^{1/2}=\left(\sum_{\alpha=0}^{k}\|f^{(\alpha)}\|_{L_2}^2\right)^{1/2},$$

则可证明$H^k(a,b)$是**完备的内积空间**，并称之为 **Sobolev 空间**。通常也用到 Sobolev 空间$H_0^k(a,b)$，它的定义是

$$H_0^k(a,b):=\{f(x)\,|\,f(x)\in H^k(a,b)\text{且}f(a)=f(b)=0\}.$$

定理5.1 $H^k(a,b)$是一个完备的内积空间，即是一个 Hilbert 空间。

证明 $H^k(a,b)$在内积$(u,v)_k$的定义下是一内积空间，这

是显然的. 下面证明 $H^k(a,b)$ 的完备性. 为此, 设 $\{f_n\}$ 是 $H^k(a,b)$ 中的Cauchy 序列, 所以只要证明存在 $f\in H^k(a,b)$, 有 $\|f_n-f\|_k\to 0\ (n\to\infty)$ 成立即可.

因为 $\{f_n\}$是 Cauchy 序列, 所以由 Sobolev 空间模的定义有

$$\|f_n-f_m\|_k^2=\sum_{\alpha=0}^{k}\|D^\alpha(f_n-f_m)\|_{L_2}^2$$
$$=\sum_{\alpha=0}^{k}\|D^\alpha f_n-D^\alpha f_m\|_{L_2}^2\to 0\quad (m,n\to\infty),$$

就有

$$\|D^\alpha f_n-D^\alpha f_m\|_{L_2}\to 0,\ \alpha=0,1,\cdots,k,$$

这表明广义微商序列$\{D^\alpha f_n\}$是 L_2 空间中的 Cauchy 序列. 由L_2 空间的完备性知: 存在$f^\alpha\in L_2(a,b)$使

$$\|D^\alpha f_n-f^\alpha\|_{L_2}\to 0,\ \alpha=0,1,\cdots,k.$$

又因为$D^\alpha f_n$ 是f_n 的广义微商, 所以据广义微商的定义有

$$\int_a^b \varphi D^\alpha f_n dx=(-1)^\alpha\int_a^b f_n D^\alpha\varphi dx,\ \alpha=0,1,\cdots,k,$$

再令$n\to\infty$, 就有

$$\int_a^b f^\alpha\varphi dx=(-1)^\alpha\int_a^b fD^\alpha\varphi dx,\ \alpha=0,1,\cdots,k.$$

最后, 据广义微商的定义, 即有

$$f^\alpha=D^\alpha f\in L_2(a,b),\ \alpha=0,1,\cdots,k,$$

故有

$$\|D^\alpha f_n-D^\alpha f\|_{L_2}\to 0,\ \alpha=0,1,\cdots,k.$$

总之, 存在 $f\in H^k(a,b)$使有

$$\|f_n-f\|_k^2=\sum_{\alpha=0}^{k}\|D^\alpha(f_n-f)\|_{L_2}^2=\sum_{\alpha=0}^{k}\|D^\alpha f_n-D^\alpha f\|_{L_2}^2\to 0$$
$$(n\to\infty),$$

即证得$H^k(a,b)$是一完备的内积空间. □

2.4 Sobolev 空间$H^k(\Omega)$和 $H_0^k(\Omega)$

对于高维区域, 有关经典导数推广为广义导数的途径以及有

关 Sobolev 空间的定义方式都可以将一维有关问题的讨论平行地推广得到。

记高维区域 $\Omega\subset R^n, \bar{\Omega}=\Omega\cup\partial\Omega$; $x=(x_1,x_2,\cdots,x_n)$, $\alpha=(\alpha_1,\alpha_2,\cdots,\alpha_n)$, $|\alpha|=\alpha_1+\alpha_2+\cdots+\alpha_n$;

$$D^{\alpha_i}u(x)=\frac{\partial^{\alpha_i}u(x)}{\partial x_i^{\alpha_i}},$$

$$D^\alpha u(x)=\frac{\partial^{\alpha_1+\alpha_2+\cdots+\alpha_n}}{\partial x_1^{\alpha_1}\partial x_2^{\alpha_2}\cdots\partial x_n^{\alpha_n}}u(x)=D^{\alpha_1}D^{\alpha_2}\cdots D^{\alpha_n}u(x).$$

在高维区域的有关问题中，**Green 公式**

$$\int_\Omega\frac{\partial u}{\partial x_i}v\,dx=-\int_\Omega\frac{\partial v}{\partial x_i}u\,dx+\int_{\partial\Omega}uv\cos(n,x_i)\,ds,\quad\forall v\in C_0^\infty(\Omega)$$

是实现变分形式和微分方程形式转换，实现微分方程弱解和经典解转换的关键，所以用Green 公式推广导数的概念。

设$u\in C^\alpha(\Omega)$，那么下列关系式显然成立：

$$\int_\Omega v\frac{\partial u}{\partial x_i}dx=-\int_\Omega u\frac{\partial v}{\partial x_i}dx,\quad\forall v\in C_0^\infty(\Omega),i=1,2,\cdots,n;$$

$$\int_\Omega v\frac{\partial^{\alpha_i}u}{\partial x_i^{\alpha_i}}dx=(-1)^{\alpha_i}\int_\Omega u\frac{\partial^{\alpha_i}v}{\partial x_i^{\alpha_i}}dx,\quad\forall v\in C_0^\infty(\Omega),$$

$$i=1,2,\cdots,n;$$

$$\int_\Omega vD^\alpha u\,dx=(-1)^{|\alpha|}\int_\Omega uD^\alpha v\,dx,\quad\forall v\in C_0^\infty(\Omega).$$

若使上面诸式在积分的意义下成立，即可把连续可微的概念推广到较弱的情形。

定义 对函数$u\in L_2(\Omega)$，如果存在$g_\alpha\in L_2(\Omega)$，使得

$$\int_\Omega vg_\alpha\,dx=(-1)^{|\alpha|}\int_\Omega uD^\alpha v\,dx,\quad\forall v\in C_0^\infty(\Omega)$$

成立，则称g_α是u的 ***α* 阶广义导数**，且记为$g_\alpha=D^\alpha u$。

直接利用广义微商的概念，可得高维区域广义微商的性质，它与一维情形广义微商的几个性质一样（略）。

仿一维 Sobolev 空间的定义，可直接定义

$$H^k(\Omega):=\Big\{f(x)\,|\,D^\alpha f\in L_2(\Omega),\ |\alpha|\leqslant k;$$

$$\|f\|_k^2:=\sum_{|\alpha|=0}^{k}\int_\Omega |D^\alpha f|^2 dx\Big\},$$

$$H_0^k(\Omega):=\{f(x)\,|\,f(x)\in H^k(\Omega),\ f(x)|_{\partial\Omega}=0\}.$$

具体地，对二维区域情形，当$f(x)\in H^2(\Omega)$时，有

$$\|f\|_0^2=\int_\Omega |f|^2 dx,$$

$$\|f\|_1^2=\int_\Omega\Big[(f)^2+\Big(\frac{\partial f}{\partial x_1}\Big)^2+\Big(\frac{\partial f}{\partial x_2}\Big)^2\Big]dx,$$

$$\|f\|_2^2=\int_\Omega\Big[(f)^2+\Big(\frac{\partial f}{\partial x_1}\Big)^2+\Big(\frac{\partial f}{\partial x_2}\Big)^2$$

$$+\Big(\frac{\partial^2 f}{\partial x_1^2}\Big)^2+\Big(\frac{\partial^2 f}{\partial x_1\partial x_2}\Big)^2+\Big(\frac{\partial^2 f}{\partial x_2^2}\Big)^2\Big]dx.$$

前面，我们利用广义导数直接定义了 Sobolev 空间$H^k(\Omega)$和$H_0^k(\Omega)$，达到了扩大函数类的目的。然而，由于广义导数和经典导数存在着直接关系，人们完全可以利用 $C^k(\Omega)$ 扩充函数类，从而可以从另一个角度研究和定义Sobolev空间。

大家知道，在$C^k(\Omega)$上定义内积和模

$$(u,v)_k:=\int_\Omega\sum_{|\alpha|=0}^{k}D^\alpha u D^\alpha v dx,$$

$$\|u\|_k^2:=(u,u)_k=\int_\Omega\sum_{|\alpha|=0}^{k}|D^\alpha u|^2 dx,$$

则 $C^k(\Omega)$ 对于内积$(\cdot,\cdot)_k$构成一个内积空间。但这个内积空间并不完备，即对任意的 Cauchy 序列$\{u_n\}\subset C^k(\Omega)$，不一定存在$\{u_n\}$的极限 $u\in C^k(\Omega)$，使$\|u_n-u\|_k\to 0$ 成立。但人们可用**函数空间完备化**的办法，对每个 Cauchy 序列的极限增补定义，使函数类$C^k(\Omega)$在$\|\cdot\|_k$的意义下完备化，并将这个经完备化的函数类记为$W^k(\Omega)$，显然，$C^k(\Omega)\subset W^k(\Omega)$，且$W^k(\Omega)$是一个完备的内积空间。

那么$H^k(\Omega)$和$W^k(\Omega)$这两个完备的内积空间有什么内在联系呢?

可以证明　若$u\in W^k(\Omega)$，则u必具有直到k阶的广义微商。

事实上，若$u\in W^k(\Omega)$，由$W^k(\Omega)$的定义，则存在$\{u_n\}\subset C^k(\Omega)$，使有

$$\|u_n-u\|_k\to 0\quad (n\to\infty),$$

且

$$\|u_n-u_p\|_k^2=\sum_{|\alpha|=0}^{k}\|D^\alpha u_n-D^\alpha u_p\|_{L_2}^2\to 0\quad (n,p\to\infty)$$

成立，从而有

$$\|D^\alpha u_n-D^\alpha u_p\|_{L_2}\to 0,\quad |\alpha|=0,1,\cdots,k,\ \ n,p\to\infty.$$

这表明$\{D^\alpha u_n\}$是$L_2(\Omega)$中的Cauchy序列，从而存在$g_\alpha\in L_2(\Omega)$使有

$$\|D^\alpha u_n-g_\alpha\|_{L_2}\to 0,\quad |\alpha|=0,1,\cdots,k,\ \ n\to\infty$$

成立。现要证明g_α就是$\{u_n\}$极限u的广义微商。因$u_n\in C^k(\Omega)$，故适合 Green 公式：

$$\int_\Omega vD^\alpha u_n dx=(-1)^{|\alpha|}\int_\Omega u_n D^\alpha v dx,\quad |\alpha|=0,1,\cdots,k,$$

$$\forall v\in C_0^\infty(\Omega);$$

再令$n\to\infty$，则有$u,u_n\in W^k(\Omega)$，使$u_n\to u$且

$$\int_\Omega vg_\alpha dx=(-1)^{|\alpha|}\int_\Omega uD^\alpha v dx,\quad |\alpha|=0,1,\cdots,k,\ \ \forall v\in C_0^\infty(\Omega).$$

据广义微商的定义可知，$g_\alpha\in L_2(\Omega)$且$g_\alpha=D^\alpha u\ (|\alpha|=0,1,\cdots,k)$，所以证得$W^k(\Omega)$中元素的确具有直到$k$阶的广义微商。

显然，$W^k(\Omega)$中元素的这一性质正说明

$$H^k(\Omega)\subset W^k(\Omega).$$

另外，还可以证明[6~7]（证明较难，略）$W^k(\Omega)\subset H^k(\Omega)$。故有

$$H^k(\Omega)=W^k(\Omega).$$

由Sobolev空间的$H^k(\Omega)$定义，很容易验证某个函数是否属于$H^k(\Omega)$，由Sobolev空间的$W^k(\Omega)$定义，又可用$\{u_n\}\subset C^k(\Omega)$来逼近和研究$u\in H^k(\Omega)$，这些都给研究和应用Sobolev空间带来很大方便。另外，由于Sobolev空间中的模表示了某种能量，例如$\int_\Omega u^2dxdy$可表示势能，$\int_\Omega(u_x^2+u_y^2)dxdy$可表示形变能，$\int_\Omega(u_{xx}^2+u_{xy}^2+u_{yy}^2)dxdy$可表示弯曲能，所以在Sobolev空间中研究数学物理问题也就相当于从能量角度研究数理问题，它更具有广泛和深刻的价值，所以Sobolev空间在应用数学范围内已广泛被应用。

§3 弱解方程的可解性

3.1 目的和任务

由于Sobolev空间的引进，人们可以在更广大的函数类和更弱的条件下研究变分问题和微分方程。

例如第一章§1中提及的弦平衡变分问题就可以在Sobolev空间中讨论，此时容许函数类和试探函数空间不再是$M=M_0=\{v\in C^2[a,b]\mid v(a)=v(l)=0\}$，而应是$\{v\in H^1\mid v(a)=v(l)=0\}$；同样地，本章§1中介绍的几个变分问题也可以在Sobolev空间中讨论之。

又如第一章§2中的两点边值问题

$$\begin{cases} Lu=-\dfrac{d}{dx}\left(p(x)\dfrac{du}{dx}\right)+q(x)u=f(x), \quad x\in(a,b), \\ u(a)=\alpha, \\ p(b)\dfrac{du(b)}{dx}+g(b)u(b)=\beta, \end{cases} \tag{5.15}$$

其相应的变分问题应描述为

$$\begin{cases} J(u) = \frac{1}{2} a(u,u) - F(u) \Rightarrow \min, \quad u \in H_E^1(a,b), \\ H_E^1(a,b) := \{u \in H^1(a,b) \mid u(a) = \alpha\}; \end{cases} \tag{5.16}$$

再施行变分法，就得到变分问题(5.16)取得唯一极小值所必须满足的条件

$$a(u,v) = F(v), \quad \forall v \in H_{0E}^1(a,b), \tag{5.17}$$

和

$$a(v,v) > 0, \quad \forall v \in H_{0E}^1(a,b), \tag{5.18}$$

式中

$$a(u,v) := \int_a^b \left(p \frac{du}{dx} \frac{dv}{dx} + quv \right) dx + g(b) u(b) v(b),$$

$$F(v) := \int_a^b fv dx + \beta v(b),$$

$$H_{0E}^1(a,b) := \{v \in H^1(a,b) \mid v(a) = 0\}.$$

对某个具体的问题，容许函数类和试探函数空间的确定是不难的。例如对边值条件

$$u(a) = \alpha, \quad u(b) = \beta,$$

应有 $H_E^1(a,b) := \{u \in H^1(a,b) \mid u(a) = \alpha, \ u(b) = \beta\}$,

$$H_{0E}^1(a,b) := \{v \in H^1(a,b) \mid v(a) = v(b) = 0\};$$

又如对边值条件

$$p(a) \frac{du(a)}{dx} + g(a) u(a) = \alpha, \quad u(b) = \beta,$$

应有 $H_E^1(a,b) := \{u \in H^1(a,b) \mid u(b) = \beta\}$,

$$H_{0E}^1(a,b) := \{v \in H^1(a,b) \mid v(b) = 0\}.$$

变分问题(5.16)及变分方程(5.17)的解与边值问题(5.15)的解是有差别和联系的。边值问题(5.15)的解在通常意义下要求 $u \in C^2(a,b)$，其近似解序列 $\{u_n\}$ 收敛于 u 是在逐点收敛的意义下，这种收敛称为**强收敛**，记为 $u_n \to u$，此时边值问题(5.15)的解 $u \in C^2(a,b)$ 称为**经典解**。但(5.16)式及(5.17)式的近似解序

列是在积分号下的收敛，是一种平均意义下的收敛，即满足$a(u_n,v)\to a(u,v)$，称为**弱收敛**，记为$u_n\rightharpoonup u$，所以(5.16)及(5.17)式的解（属于H^1）称为**弱解**。(5.17)式又称为(5.15)式的**弱解方程**。只有$u\in C^2(a,b)$时，经典解和弱解才是一致的。

弱解方程(5.17)的解是变分问题(5.16)的驻点解；要变分问题取得**整体极小解**，二次泛函$a(\cdot,\cdot)$必须在试探函数空间中满足**正定性条件**(5.18)，这种正定性条件在Sobolev空间中的确切描述是

定义 称二次泛函$a(\cdot,\cdot)$是正定的，若存在实数γ，使有

$$a(v,v)\geqslant\gamma^2\|v\|_1^2,\quad \forall v\in H_{0E}^1$$

成立。

上述关系在高维区域情形也同样适用，因此，要研究变分问题的解（或弱解）的性质，就必须研究二次泛函$a(\cdot,\cdot)$的性质及弱解方程的可解性。

3.2 一维问题中$a(\cdot,\cdot)$的性质

先介绍两个重要的不等式。

若$u\in H^1(a,b)$且$u(a)=0$，则有 **Friedrichs 不等式**：

$$|u|_1^2\geqslant\frac{2}{(b-a)^2}\|u\|_0^2. \tag{5.19a}$$

证明 因为对$\forall u\in H^1(a,b)$且$u(a)=0$，都有

$$u(x)=u(x)-u(0)=\int_a^x u'(t)dt,$$

利用Schwarz 不等式就有

$$|u(x)|^2=\left|\int_a^x 1\cdot u'(t)dt\right|^2\leqslant\int_a^x|u'(t)|^2dt\cdot\int_a^x dt$$

$$\leqslant(x-a)\int_a^b|u'(t)|^2dt,$$

两边积分得

$$\int_a^b|u(x)|^2dx\leqslant\frac{1}{2}(b-a)^2\int_a^b|u'(t)|^2dt,$$

整理后得(5.19a)。 □

若$u\in H^1(a,b)$，则有Sobolev 不等式：

$$|u(x)|\leqslant c\|u\|_1, c=\sqrt{\frac{2}{(b-a)}+\frac{2(b-a)}{3}}, \forall x\in[a,b] \tag{5.19b}$$

证明 为证明需要，构造函数$f(x)\in H^1(a,b)$，

$$f(x)=\begin{cases}2(x-a)/(b-a), & a\leqslant x\leqslant a+(b-a)/2,\\ 1, & a+(b-a)/2\leqslant x\leqslant b;\end{cases}$$

并在$x\in[a,(b-a)/2]$时考虑辅助函数$w=uf$。
显然$w(a)=0$，且

$$u(x)=w(x)-w(a)=\int_a^x w'(t)\,dt=\int_a^x(u'f+uf')\,dt.$$

再用三角不等式和Schwarz 不等式，就有

$$|u(x)|\leqslant\left|\int_a^b u'f dx\right|+\left|\int_a^b uf'dx\right|$$

$$\leqslant\sqrt{\frac{2(b-a)}{3}}\left(\int_a^b|u'|^2dx\right)^{1/2}+\sqrt{\frac{2}{(b-a)}}\left(\int_a^b|u|^2dx\right)^{1/2}$$

$$\leqslant\sqrt{\frac{2}{(b-a)}+\frac{2(b-a)}{3}}\|u\|_1.$$

改变$f(x)$的作法，类似可证$x\in[(b-a)/2,b]$的情形。故证得(5.19b)式成立。 □

下面就一维边值问题(5.15)讨论其对应的弱解方程(5.17)中的$a(\cdot,\cdot)$的性质。

定理5.2 若$p(x)\in C^0$且$p_{\min}>0$；$q(x)\in C^0$且$q_{\min}\geqslant 0$；$g(x)\in C^0$且$g_{\min}\geqslant 0$；则二次泛函

$$a(u,v)=\int_a^b(pu'u'+quv)\,dx+g(b)u(b)v(b)$$

有如下性质：

1) $a(u,v)$是双线性对称的二次泛函；

2）有界性：

$$|a(u,v)|\leqslant M\|u\|_1\|v\|_1,\quad \forall u,v\in H^1(a,b);$$

3）正定性：

$$|a(v,v)|\geqslant \gamma^2\|v\|_1^2,\ \gamma\in R,\ \forall v\in H_{0E}^1(a,b)$$

式中 $H_{0E}^1:=\{v|v\in H^1(a,b),v(a)=0\}$。

证明 利用双线性泛函的定义

$$a(u_1+u_2,v)=a(u_1,v)+a(u_2,v),$$

$$a(u,v_1+v_2)=a(u,v_1)+a(u,v_2),$$

和对称泛函的定义$a(u,v)=a(v,u)$，易知1）显然成立。

利用Schwarz 不等式和Sobolev 不等式(5.19b)就有

$$\begin{aligned}|a(u,v)|&\leqslant\left|\int_a^b(pu'v'+quv)dx\right|+|g(b)u(b)v(b)|\\&\leqslant p_{max}\|u'\|_0\|v'\|_0+q_{max}\|u\|_0\|v\|_0+\widetilde{M}\|u\|_1\|v\|_1\\&\leqslant M\|u\|_1\|v\|_1.\end{aligned}$$

所以性质2）成立。

最后，利用 Friedrichs 不等式(5.19a)，有

$$\begin{aligned}|a(v,v)|&=\left|\int_a^b[p(v')^2+qv^2]dx+g(b)v^2(b)\right|\\&\geqslant\left|\int_a^b p(v')^2dx\right|\quad(\because q_{min}\geqslant0,\ g_{min}\geqslant0)\\&\geqslant\frac{1}{2}p_{min}[|v|_1^2+|v|_1^2]\\&\geqslant\frac{1}{2}p_{min}\left[|v|_1^2+\frac{2}{(b-a)^2}\|v\|_0^2\right]\\&\geqslant\gamma^2\|v\|_1^2.\end{aligned}$$

就证得性质3）. □

顺便指出，利用 Friedrichs 不等式和正定性条件即知，在 $H_{0E}^1=\{v|v\in H^1(a,b),\ v(a)=0\}$ 函数空间中，半模 $|v|_1=\left(\int_a^b(v')^2dx\right)^{1/2}$ 与**全模**$\|v\|_1$是**等价**的，即存在与 v 无关的常数c_1

和c_2使下式成立：

$$c_1\|v\|_1\leqslant|v|_1\leqslant c_2\|v\|_1,\quad \forall v\in H^1_{0E}(a,b). \tag{5.20}$$

3.3 二维问题中$a(\cdot,\cdot)$的性质

在二维椭圆边值问题中，最有广泛代表性的是

$$\begin{cases} Lu=-\nabla(p\nabla u)+ru=f,\ (x,y)\in\Omega, \\ u(x,y)=\alpha(x,y),\ (x,y)\in\Gamma_1, \\ p\partial u/\partial n+\sigma u=\beta,\ (x,y)\in\Gamma_2, \end{cases} \tag{5.21}$$

其中，$\partial\Omega=\Gamma_1\cup\Gamma_2$。与其相应的变分问题和弱解方程分别是

$$J(u)=\frac{1}{2}a(u,u)-F(u)\Rightarrow\min,\quad u\in H^1_E(\Omega),$$

$$a(u,v)=F(v),\quad u\in H^1_E(\Omega),\quad \forall v\in H^1_{0E}(\Omega), \tag{5.22}$$

式中

$$H^1_E(\Omega):=\{u\in H^1(\Omega)\,|\,u|_{\Gamma_1}=\alpha\},$$
$$H^1_{0E}(\Omega):=\{v\in H^1(\Omega)\,|\,v|_{\Gamma_1}=0\},$$

$$a(u,v):=\int_\Omega[p\nabla u\nabla v+ruv]\,dxdy+\int_{\Gamma_2}\sigma uvds,$$

$$F(v):=\int_\Omega fvdxdy+\int_{\Gamma_2}\beta vds.$$

为了证明二次泛函$a(\cdot,\cdot)$的有界性与正定性，下面先介绍**Sobolev 迹定理**和**Friedrichs 不等式**。

定理5.3 若$u\in H^1(\Omega)$，在区域$\bar{\Omega}$中的每一条光滑的曲线l上（特别地，l可以是边界Γ），函数u作为l上的一元函数(称为函数的迹）是平方可积的，则有不等式

$$\left(\int_l u^2ds\right)^{1/2}\leqslant c\|u\|_1 \tag{5.23}$$

成立，其中常数c与u无关。

定理5.4 若$u\in H^1(\Omega)$且在Ω的部分边界Γ_1上满足本质边界条件$u=0$，则**Friedrichs不等式**

$$\left(\int_\Omega u^2 dxdy\right)^{1/2} \leqslant c\left\{\int_\Omega\left[\left(\frac{\partial u}{\partial x}\right)^2+\left(\frac{\partial u}{\partial y}\right)^2\right]dxdy\right\}^{1/2} \qquad (5.24)$$

成立，其中常数 c 与 u 无关。

证明 不失一般性，设区域 $\overline{\Omega}$ 完全包含在中心于坐标原点且边长为 $2R$ 的正方形中，在 $\overline{\Omega}$ 外 $u\equiv 0$，并设 Γ_1 的纵坐标 $\alpha\leqslant y\leqslant\beta$，如图5.4示。

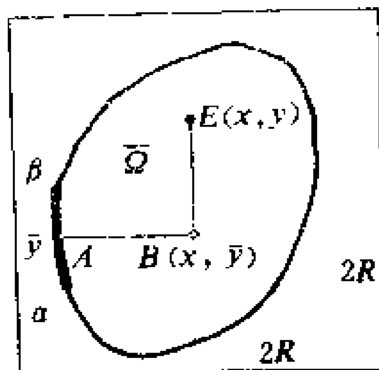


图5.4

过 Γ_1 上纵坐标为 $\bar{y}$ 的 A 点作平行于 x 轴的直线段 AB，沿 AB 积分，有

$$u(B)=u(B)-u(A)=\int_A^B \frac{\partial u(\xi,\bar{y})}{\partial \xi}d\xi.$$

又由 Schwarz 不等式，有

$$|u(x,\bar{y})|^2=\left|\int_A^B \frac{\partial u(\xi,\bar{y})}{\partial \xi}d\xi\right|^2 \leqslant 2R\int_A^B\left|\frac{\partial u(\xi,\bar{y})}{\partial \xi}\right|^2 d\xi.$$

将上式两边对 $\bar{y}$ 积分($\alpha\leqslant\bar{y}\leqslant\beta$)，就有

$$\int_\alpha^\beta |u(x,\bar{y})|^2 d\bar{y}=\int_\alpha^\beta |u(x,y)|^2 dy$$

$$\leqslant 2R\int_A^B\int_\alpha^\beta\left|\frac{\partial u(\xi,\bar{y})}{\partial \xi}\right|^2 d\bar{y}d\xi \leqslant 2R\int_A^B\int_{-R}^{+R}\left|\frac{\partial u(\xi,y)}{\partial \xi}\right|^2 dyd\xi$$

$$\leqslant 2R\int_\Omega\left|\frac{\partial u(x,y)}{\partial x}\right|^2 dxdy. \qquad (5.25)$$

若区域 $\overline{\Omega}$ 含于带形域 $\alpha\leqslant y\leqslant\beta$ 中，则将(5.25)式再对 x 积分就可证得(5.24)式。否则，还要过 $B(x,\bar{y})$ 作平行于 y 轴的直线段 BE，并沿 BE 积分，有

$$u(x,y)=u(E)=u(B)+\int_{\bar{y}}^{y}\frac{\partial u(x,\eta)}{\partial \eta}d\eta.$$

利用Cauchy 不等式和Schwarz 不等式，就有

$$|u(x,y)|^2=\left[u(x,\bar{y})+\int_{\bar{y}}^{y}\frac{\partial u(x,\eta)}{\partial\eta}d\eta\right]^2$$

$$\leqslant|u(x,\bar{y})|^2+\left|\int_{\bar{y}}^{y}\frac{\partial u(x,\eta)}{\partial\eta}d\eta\right|^2+2\left|u(x,\bar{y})\int_{\bar{y}}^{y}\frac{\partial u(x,\eta)}{\partial\eta}d\eta\right|$$

$$\leqslant2\left[|u(x,\bar{y})|^2+\left|\int_{\bar{y}}^{y}\frac{\partial u(x,\eta)}{\partial\eta}d\eta\right|^2\right]$$

$$\leqslant2|u(x,\bar{y})|^2+2R\int_{-R}^{+R}\left|\frac{\partial u(x,\eta)}{\partial\eta}\right|^2d\eta,$$

两边再对 $\bar{y}$ 积分

$$|u(x,y)|^2(\beta-\alpha)\leqslant2\int_{\alpha}^{\beta}|u(x,\bar{y})|^2d\bar{y}+(2R)^2\int_{-R}^{+R}\left|\frac{\partial u(x,y)}{\partial y}\right|^2dy\ ;$$

还利用刚才得到的(5.25)式，就有

$$|u(x,y)|^2(\beta-\alpha)\leqslant4R\int_{\Omega}\left(\frac{\partial u}{\partial x}\right)^2dxdy+4R^2\int_{-R}^{+R}\left(\frac{\partial u}{\partial y}\right)^2dy\ .$$

最后对x和y积分，就证得(5.24)式。□

下面证明二维问题中二次泛函$a(\cdot,\cdot)$的性质。

定理5.5 若 $p\in C^0(\bar{\Omega})$ 且$p_{\min}>0$；$r\in C^0(\bar{\Omega})$且$r_{\min}\geqslant0$；$\sigma\in C^0(\Gamma_2)$且$\sigma_{\min}\geqslant0$；则二次泛函

$$a(u,v)=\int_{\Omega}[p\nabla u\nabla v+ruv]dxdy+\int_{\Gamma_2}\sigma uvds$$

适合下述性质：

1) $a(u,v)$是双线性对称泛函；

2) 有界性 $|a(u,v)|\leqslant M\|u\|_1\|v\|_1,\quad\forall u,v\in H^1(\Omega)$；

3) 正定性 $|a(v,v)|\geqslant r^2\|v\|_1^2,\quad\forall v\in H^1_{0E}(\Omega)$，

式中，$H^1_{0E}(\Omega)=\{v\in H^1(\Omega)\,|\,v|_{\Gamma_1}=0\}$。

证明 性质 1) 显然，现证有界性和正定性。

利用 Sobolev 迹定理，就有

$$\left|\iint_\Omega [p\nabla u\nabla v + ruv]dxdy + \int_{\Gamma_2}\sigma uv\, ds\right|$$

$$\leqslant \max(|p_{max}|, |r_{max}|)\left|\iint_\Omega (\nabla u\nabla v + uv)dxdy\right|$$

$$+ |\sigma_{max}|\left|\int_{\Gamma_2} uvds\right|$$

$$\leqslant c_1(\|\nabla u\|_0\|\nabla v\|_0 + \|u\|_0\|v\|_0) + c_2\|u\|_1\|v\|_1$$

$$\leqslant c\|u\|_1\|v\|_1,$$

性质2）得证。最后证性质3）。

利用Friedrichs不等式，就有

$$|a(v,v)| = \left|\iint_\Omega [p\nabla u\nabla v + ruv]dxdy + \int_{\Gamma_2}\sigma v\, vds\right|$$

$$\geqslant \left| p_{min}\int_\Omega \nabla v\nabla v dxdy\right| \quad (r\geqslant 0, \sigma\geqslant 0)$$

$$\geqslant c\int_\Omega (|\nabla v|^2 + |v|^2)dxdy$$

$$= \gamma^2\|v\|_1^2.$$

就证得$a(v,v)$在试探函数空间$H^1_{0E}(\Omega)$中正定。 □

顺便指出，利用Sobolev迹定理也很容易证得弱解方程右端项$F(v)$的有界性。

3.4 弱解方程的可解性

大家知道，从模线性空间X的子集D到模线性空间Y的映象T： $x\to y = T(x)$叫作**算子**；D称为该算子的**定义域**，记着$D(T)$；$R(T) = \{y|y = T(x),\ x\in D(T)\}$称作该算子的**值域**。

对于线性算子，连续性和有界性是等价的。

对有界线性算子T，定义其模为

$$\|T\|: = \sup_{\substack{x\in D(T)\\ \|x\|\neq 0}} \frac{\|T(x)\|_Y}{\|x\|_X} = \sup_{\substack{x\in D(T)\\ \|x\| = 1}} \|T(x)\|_Y;$$

由于在实数域上模都与绝对值$|\cdot|$等价，故对有界线性泛函F有

$$\|F\| := \sup_{\substack{x \in D(F) \\ \|x\| = 1}} |F(x)|; \tag{5.26}$$

因而有

$$\|T(x)\|_Y \leqslant \|T\| \|x\|_X, \quad |F(x)| \leqslant \|F\| \|x\|_X.$$

模线性空间X上的一切有界线性泛函组成的集合也构成一个线性空间，用(5.26)式定义模，则此集合也是模线性空间，通常称为**共轭空间**，并记着X'.

如果线性算子T从$D(T)$到$R(T)$的映象是一一对应的，则作为映象T：$x \to Tx$的逆映象T^{-1}具有性质：

1) $T^{-1}(T(x)) = x, \quad \forall x \in D(T)$,

2) $T(T^{-1}(y)) = y, \quad \forall y \in R(T)$,

3) $D(T^{-1}) = R(T), \quad R(T^{-1}) = D(T)$.

显然T^{-1}也是线性算子，它称为T的**逆算子**.

有界逆定理 线性算子T具有有界逆算子的充要条件是T下有界，即$\exists\, a > 0$使有

$$\|T(x)\| \geqslant a\|x\|, \quad \forall x \in D(T).$$

证明 必要性 因T^{-1}有界，故存在$M > 0$使有

$$\|x\| = \|T^{-1}(T(x))\| \leqslant M\|T(x)\|, \quad \forall x \in D(T),$$

选$a = 1/M$即证得必要性.

充分性 因$\|T(x)\| \geqslant a\|x\|$，由$T(x) = 0$即有$x = 0$，所以$T^{-1}$存在．再证$T^{-1}$是有界算子，事实上，对$\forall y \in D(T^{-1}) = R(T)$，$\exists x \in D(T)$，使$y = T(x)$，所以

$$\|T^{-1}(y)\| = \|T^{-1}(T(x))\| = \|x\| \leqslant \frac{1}{a}\|T(x)\| = \frac{1}{a}\|y\|,$$

即知T^{-1}有界，且$\|T^{-1}\| \leqslant \dfrac{1}{a}$. □

内积空间和一般模线性空间的重要区别，在于其中可以引入“直交”的概念．若记内积为$[\cdot, \cdot]$，则当$[x, y] = 0$时就称内积空间H中的两个元素x和y是直交的．设M是H中的一个子集，

H中那些与M的一切元素直交的元素的全体称为M的**直交补**，记着$M^\perp$，显然$M^\perp$是H的子空间。

直交分解定理 设M是 Hilbert 空间H中的一个闭子空间，则对$\forall x\in H$，必存在唯一的$x_0\in M$，使$z=x-x_0\in M^\perp$且x可写成**直交和**形式，即$x=x_0+z$。

在Hilbert 空间H中，固定某元素x，则内积$[x,y]$确定了一个关于y的线性泛函$F(y)$且有

$$F(y)=[x,y],\quad \forall y\in H;$$

反之任给一个定义在H上的有界线性泛函$F(y)$，是否一定存在元素$x_F\in H$，使有$F(y)=[x_F,y]$呢？回答是肯定的。

Riesz 表现定理 设H是Hilbert 空间，则对$\forall F\in H'$，必存在唯一的$x_F\in H$，使

$$F(y)=[x_F,y],\quad \forall y\in H. \tag{5.27}$$

证明 记F的零空间$N(F)=\{y\mid y\in H且F(y)=0\}$，易知$N(F)$是$H$的闭子空间。

若$N(F)=H$，则可取$x_F=0$，就有(5.27)式成立。

若$N(F)\neq H$，则必存在$x\in H\setminus N(F)$。由直交分解定理，对于$x_0\in N(F)$有$y_0=x-x_0\in N(F)^\perp$，且$F(g_0)\neq 0$，于是

$$y-\frac{F(y)}{F(y_0)}y_0\in N(F),\quad \forall y\in H,$$

从而利用$[\cdot,\cdot]$的定义有

$$\left[y-\frac{F(y)}{F(y_0)}y_0,y_0\right]=[y,y_0]-\frac{F(y)}{F(y_0)}[y_0,y_0],$$

即

$$F(y)=\frac{F(y_0)}{[y_0,y_0]}[y,y_0]=\left[y,\ \frac{F(y_0)}{[y_0,y_0]}y_0\right],$$

所以取$x_F=\dfrac{F(y_0)}{[y_0,y_0]}y_0$即可证得(5.27)式成立。

唯一性证明是容易的，事实上若设$\bar{x}_F$也使(5.27)式成立，则

有$[\bar{x}_F - x_F, y] = 0$ ($\forall y \in H$)，取$y = \bar{x}_F - x_F$，即得 $\bar{x}_F = x_F$。□

下面讨论本段的主要问题，即弱解方程的可解性。讨论这个问题不仅是理论研究弱解的需要也是把握有限元方程是否唯一可解的依据。

定理5.6 (Lax-Milgram) 设H是Hilbert空间，$a(u,v)$是H上的双线性泛函，且满足

$$|a(u,v)| \leqslant M\|u\|\|v\|, \quad |a(v,v)| \geqslant \gamma^2\|v\|^2,$$

其中$\|\cdot\|$表示H上的模，M、γ^2都是正常数；设$F(v)$是H上的有界线性泛函；则存在唯一的$u_0 \in H$，使得

$$a(u_0,v) = F(v), \quad \forall v \in H$$

成立，即弱解方程$a(u,v) = F(v)$在H中是唯一可解的。

证明 先考虑$a(u,v)$是对称的。由于$a(u,v)$具有对称、正定的双线性形式，所以它可以作为空间H的新内积（满足内积的诸条件），并记着$[u,v]$。又因为$a(\cdot,\cdot)$的有界性和正定性，于是由

$$\gamma^2\|v\|^2 \leqslant [v,v] \leqslant M\|v\|^2$$

知新内积$[\cdot,\cdot]$产生的模与H中原来的模$\|\cdot\|$等价，从而$F(v)$对于新模$[v,v]^{\frac{1}{2}}$仍是有界的线性泛函。故由Riesz表现定理知，存在唯一的$u_0 \in H$，使

$$F(v) = [u_0,v], \quad \forall v \in H,$$

亦即

$$a(u_0,v) = F(v), \quad \forall v \in H$$

成立，这就证明了弱解方程是唯一可解的。

再考虑$a(u,v)$是不对称的情形。因$a(u,v)$非对称，它不能再当作内积。但对任一个固定的$u \in H$，$a(u,v)$是在H上的关于v的有界线性泛函，故由Riesz表现定理，存在唯一元素$Tu \in H$（它显然依赖于u，所以记为Tu）使有

$$a(u,v) = (Tu,v), \quad \forall v \in H,$$

其中，$(\cdot,\cdot)$也表示H上的内积。

显然，T是H上的线性算子；若令$v=Tu\in H$，于是由

$$\|Tu\|^2=(Tu,Tu)=a(u,Tu)\leqslant M\|u\|\|Tu\|,$$
$$\|Tu\|\leqslant M\|u\|$$

知T是有界线性算子；又由

$$\gamma^2\|v\|^2\leqslant a(v,v)=(Tv,v)\leqslant\|Tv\|\|v\|,$$
$$\gamma^2\|v\|\leqslant\|Tv\|,$$

知T下有界。故根据有界逆定理，T^{-1}存在且上有界，$\|T^{-1}\|\leqslant\frac{1}{\gamma^2}$。

若T的值域是整个空间H，即$R(T)=H$，则T^{-1}是H上的有界线性算子。而由Riesz表现定理，对$F\in H'$存在$g\in H$使

$$F(v)=(g,v)=a(T^{-1}g,v),\quad\forall v\in H$$

成立，所以$T^{-1}g=u_0$就是弱解方程在H中的唯一解。

下面再进一步证明$R(T)=H$。显然$R(T)$是H的子空间，我们说它还是闭子空间，事实上，设$\|Tx_n-Tx_m\|\to0$，则由

$$\|x_n-x_m\|\leqslant\frac{1}{\gamma^2}\|Tx_n-Tx_m\|$$

知$\|x_n-x_m\|\to0$，故存在$x_0\in H$使有$x_n\to x_0$；又由

$$\|Tx_n-Tx_0\|\leqslant M\|x_n-x_0\|$$

知$Tx_n\to Tx_0$，即$R(T)$是闭集，$R(T)$是H的闭子空间。如果假设$R(T)\neq H$，则由直交分解定理，必存在$w_0\neq0$，$w_0\in R(T)^{\perp}$；从而有

$$a(w_0,w_0)=(Tw_0,w_0)=0,$$

即知$w_0=0$；与假设矛盾，所以反证得$R(T)=H$。 □

利用Lax-milgram定理立即可知，与二阶椭圆边值问题(5.15)和(5.21)相应的弱解方程(5.17)和(5.22)是唯一可解的，当然与(5.17)和(5.22)式相应的有限元方程也是唯一可解的。

§4 线性元误差估计

4.1 有限元解的一个性质

由第一、二章可知，为了用有限元方法求弱解方程

$$a(u,v)=F(v),\quad \forall v\in H^1_{0E} \tag{5.28}$$

(H^1_{0E}是试探函数空间，$a(\cdot,\cdot)$和$F(\cdot)$满足Lax-milgram定理中的条件)，需构造分段(片)可微的有限维函数空间$V_h\subset H^1_{0E}$，再利用Ritz-Galerkin 方法在V_h中求满足

$$a(u_h,v_h)=F(v_h),\quad \forall v_h\in V_h \tag{5.29}$$

的有限元解u_h。

下面定理讨论了有限元解的性质。

定理5.7 设u是弱解方程(5.28)的解，u_h是有限元方程(5.29)的解，则

$$a(u-u_h,v_h)=0,\quad \forall v_h\in V_h, \tag{5.30}$$

$$\|u-u_h\|_1\leqslant\beta\inf\|u-v_h\|_1,\quad \forall v_h\in V_h, \tag{5.31}$$

其中β是常数。

证明 因u和u_h依次满足(5.28)和(5.29)式，且因$v_h\in V_h\subset H^1_{0E}$，所以由(5.28)式有

$$a(u,v_h)=F(v_h),\quad \forall v_h\in V_h, \tag{5.32}$$

故由(5.29)和(5.32)两式相减，即得(5.30)式。

利用$a(u,v)$在H^1中的有界性：

$$|a(u,v)|\leqslant M\|u\|_1\|v\|_1,\quad \forall u,v\in H^1,$$

和$a(v,v)$在试探函数空间H^1_{0E}中的正定性：

$$\gamma^2\|v\|_1^2\leqslant a(v,v),\quad \forall v\in H^1_{0E},$$

且利用(5.30)式，即有

$$\gamma^2\|u-u_h\|_1^2\leqslant a(u-u_h,u-u_h)$$

$$=a(u-u_h,u)-a(u-u_h,u_h)=a(u-u_h,u)-a(u-u_h,v_h)$$

$$= a(u-u_h, u-v_h) \leqslant M\|u-u_h\|_1\|u-v_h\|_1, \quad \forall v_h \in V_h,$$

即

$$\|u-u_h\|_1 \leqslant \beta\|u-v_h\|, \quad \forall v_h \in V_h, \ \beta=(M/\gamma^2)^{\frac{1}{2}}$$

注意到 v_h 的任意性，所以有

$$\|u-u_h\|_1 \leqslant \beta \inf\|u-v_h\|_1, \quad \forall v_h \in V_h.$$

这就证得定理5.7. □

定理 5.7 不仅给出了抽象的有限元误差估计，而且深刻地揭示了有限元解（或 Ritz-Galerkin 方程的解）的性质。参见图5.5。其一，由$a(\cdot,\cdot)$的对称、正定和双线性性质，$a(u,v)$可看作一种新的内积。(5.30)式表明 $u-u_h$ 与V_h中任一元素 v_h 的内积为零（正交），所以u_h 是 u 在V_h中的投影；或曰在$a(\cdot,\cdot)$意义下，u_h是V_h中的极小解（广义**最小二乘解**）。其二，因$\|\cdot\|_1$表明了 Sobolev 空间里的元素间的距离，所以(5.31)式表明，在$\|\cdot\|_1$尺度下，u 到 u_h 的距离是 u 到 V_h 中任一元素距离中最小的一个；或曰，u_h 是在$\|\cdot\|_1$意义下的**最佳逼近解**。其三，当V_h的维数 $n\to\infty$时，即有限元网格参数$h\to 0$时，$V_h\to H^1$，从而$\inf\|u-v_h\|_1\to 0$，由(5.31)式即知$\|u-u_h\|_1\to 0$，所以有限元解u_h收敛于弱解。

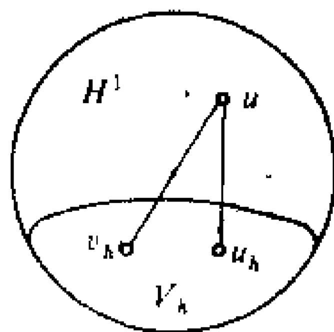


图5.5

4.2 两点边值问题线性元的误差估计

为简单起见，考虑边值问题

$$\begin{cases} -\dfrac{d}{dx}\left(p\dfrac{du}{dx}\right)+qz=f, & (x,y)\in(a,b), \\ u(a)=0, \\ u(b)=0 \text{ 或 } u'(b)=0 \end{cases} \tag{5.33}$$

的线性有限元解的误差估计。与其对应的弱解方程和有限元方程是

$$a(u,v)=F(v),\quad \forall v\in H^1_{0E}(a,b), \tag{5.34}$$

和

$$a(u_h,v_h)=F(v_h),\quad \forall v_h\in V_h, \tag{5.35}$$

这里的记号同于第一章，不累述。

A.H^1 模的估计

由定理5.7知，在H^1模度量下，有限元解的误差估计满足不等式(5.31)。这就是说，对u_h是线性有限元解，V_h是分段线性的有限维函数空间($h=\max h_i$)，将v_h特别地取为u的**线性插值函数**$u_I\in V_h$来说，(5.31)式仍然成立，即有

$$\|u-u_h\|_1\leqslant\beta\|u-u_I\|_1$$

成立。从而，线性元H^1模估计归结为**线性插值误差**的H^1模估计。下面给出一维二阶椭圆边值问题线性有限元解的误差估计。

定理5.8 设$u\in C^2[a,b]$是(5.33)或(5.34)式的解，u_h是其线性有限元解，则有估计

$$\|u-u_h\|_1\leqslant ch\|u''\|_0. \tag{5.36}$$

证明 在任一单元$I_i=[x_{i-1},x_i]$内考虑误差

$$e(x)=u(x)-u_I(x),$$

显然$e(x_{i-1})=e(x_i)=0$。用仿射变换$\xi=x-x_{i-1}$把单元$[x_{i-1},x_i]$变到$[0,1]$，并在$[0,1]$上将$e(x)$展成Fourier级数：

$$\begin{cases} e(x_{i-1}+h_i\xi)=\sum\limits_{n=1}^{\infty}a_n\sin n\pi\xi, \\ a_n=2\int_0^1 e(x_{i-1}+h_it)\sin n\pi t\,dt,\quad n=1,2,\cdots \end{cases} \tag{5.37}$$

由 Parseval 等式：

$$\int_{x_{i-1}}^{x_i}|e(x)|^2dx=h_i\int_0^1|e(x_{i-1}+h_i\xi)|^2d\xi=\frac{h_i}{2}\sum_{n=1}^{\infty}a_n^2,$$

而

$$a_n=2\int_0^1 e(x_{i-1}+h_it)\sin n\pi t\,dt$$

$$=\frac{2h_i}{n\pi}\int_0^1 e'(x_{i-1}+h_it)\cos n\pi t\,dt$$

$$=-\frac{2h_i^2}{(n\pi)^2}\int_0^1 e''(x_{i-1}+h_it)\sin n\pi t\,dt$$

$$=-\frac{2h_i}{(n\pi)^2}\int_{x_{i-1}}^{x_i} e''(x)\sin n\pi\left(\frac{x-x_{i-1}}{h_i}\right)dx,$$

$$|a_n|\leqslant\frac{2h_i}{(n\pi)^2}\int_{x_{i-1}}^{x_i}|e''(x)|dx$$

$$\leqslant\frac{2h_i^{3/2}}{(n\pi)^2}\left[\int_{x_{i-1}}^{x_i}|e''(x)|^2dx\right]^{1/2},$$

$$|a_n|^2\leqslant\frac{4h_i^3}{(n\pi)^4}\int_{x_{i-1}}^{x_i}|u''(x)|^2dx,$$

所以

$$\int_{x_{i-1}}^{x_i}|e(x)|^2dx=\frac{h_i}{2}\sum_{n=1}^{\infty}a_n^2$$

$$\leqslant\frac{h_i}{2}\cdot\frac{4h_i^3}{\pi^4}\int_{x_{i-1}}^{x_i}|u''(x)|^2dx\cdot\left[\sum_{n=1}^{\infty}\frac{1}{n^4}\right]$$

$$\leqslant ch_i^4\int_{x_{i-1}}^{x_i}|u''(x)|^2dx,$$

从而有

$$\int_a^b|e(x)|^2dx=\sum_i\int_{x_{i-1}}^{x_i}|e(x)|^2dx$$

$$\leqslant c_1h^4\sum_i\int_{x_{i-1}}^{x_i}|u''(x)|^2dx=c_1h^4\int_a^b|u''(x)|^2dx$$

即

$$\|e(x)\|_0\leqslant c_1h^2\|u''(x)\|_0. \tag{5.38}$$

下面再估计$\|e'(x)\|_0$，由(5.37)式和Parsevel 等式有

$$\int_{x_{i-1}}^{x_i}|e'(x)|^2dx=\frac{h_i}{2}\sum_{n=1}^{\infty}\left(\frac{n\pi}{h_i}\right)^2a_n^2,$$

和

$$\int_{x_{i-1}}^{x_i} |e''(x)|^2 dx = \frac{h_i}{2}\sum_{n=1}^{\infty}\left(\frac{n\pi}{h_i}\right)^4 a_n^2.$$

将这两个等式的右端项逐项比较，显见

$$\left(\frac{n\pi}{h_i}\right)^2 a_n^2 \leqslant \left(\frac{h_i}{\pi}\right)^2\left(\frac{n\pi}{h_i}\right)^4 a_n^2, \quad n=1,2,\cdots$$

成立，因此有

$$\int_{x_{i-1}}^{x_i} |e'(x)|^2 dx \leqslant \left(\frac{h_i}{\pi}\right)^2 \int_{x_{i-1}}^{x_i} |e''(x)|^2 dx$$

$$=\left(\frac{h_i}{\pi}\right)^2 \int_{x_{i-1}}^{x_i} |u''(x)|^2 dx \leqslant \left(\frac{h}{\pi}\right)^2 \int_{x_{i-1}}^{x_i} |u''(x)|^2 dx.$$

则得到关于插值误差的一阶导数的估计：

$$\|e'(x)\|_0 \leqslant c_2 h \|u''(x)\|_0. \tag{5.39}$$

最后，利用(5.38)和(5.39)式即有

$$\|u-u_I\|_1 = \left[\int_a^b |e(x)|^2 dx + \int_a^b |e'(x)|^2 dx\right]^{\frac{1}{2}}$$

$$\leqslant \|e(x)\|_0 + \|e'(x)\|_0 \leqslant ch\|u''(x)\|_0,$$

从而(5.36)式成立。 □

补充：Parseval 等式的证明.

设$f(x)(x\in[-\pi,\pi], f(-\pi)=f(\pi)=0)$可展开为Fourier级数，即

$$f(x) = \frac{a_0}{2} + \sum_{k=1}^{\infty}(a_k\cos kx + b_k\sin kx),$$

$$a_k = \frac{1}{\pi}\int_{-\pi}^{+\pi} f(x)\cos kx\,dx,$$

$$b_k = \frac{1}{\pi}\int_{-\pi}^{+\pi} f(x)\sin kx\,dx,$$

则 Parseval 等式

$$\frac{1}{\pi}\int_{-\pi}^{\pi} |f(x)|^2 dx = \frac{a_0^2}{2} + \sum_{k=1}^{\infty}(a_k^2 + b_k^2) \tag{5.40}$$

成立。

事实上，若考虑$f(x)$和

$$S_n=\frac{\alpha_0}{2}+\sum_{k=1}^{n}(\alpha_k\cos kx+\beta_k\sin kx)$$

的**平方均值误差**δ_n^2，即有

$$\delta_n^2=\frac{1}{2\pi}\int_{-\pi}^{+\pi}\left[f(x)-\frac{\alpha_0}{2}-\sum_{k=1}^{n}(\alpha k\cos kx+\beta_k\sin kx)\right]^2dx,$$

利用和的平方公式就得

$$\delta_n^2=\frac{1}{2\pi}\int_{-\pi}^{\pi}|f(x)|^2dx-\frac{\alpha_0}{2\pi}\int_{-\pi}^{\pi}f(x)dx$$

$$-\frac{1}{\pi}\sum_{k=1}^{n}\left[\alpha_k\int_{-\pi}^{+\pi}f(x)\cos kxdx+\beta_k\int_{-\pi}^{\pi}f(x)\sin kxdx\right]$$

$$+\frac{\alpha_0^2}{4}+\frac{1}{2}\sum_{k=1}^{n}(\alpha_k^2+\beta_k^2)$$

$$=\frac{1}{2\pi}\int_{-\pi}^{\pi}|f(x)|^2dx-\frac{\alpha_0a_0}{2}-\sum_{k=1}^{n}(\alpha_ka_k+\beta_kb_k)$$

$$+\frac{\alpha_0}{4}+\frac{1}{2}\sum_{k=1}^{n}(\alpha_k^2+\beta_k^2),$$

加减下面的和

$$\frac{a_0^2}{4}+\frac{1}{2}\sum_{k=1}^{n}(a_k^2+b_k^2),$$

就有

$$\delta_n^2=\frac{1}{2\pi}\int_{-\pi}^{\pi}|f(x)|^2dx-\frac{a_0^2}{4}-\frac{1}{2}\sum_{k=1}^{n}(a_k^2+b_k^2)$$

$$+\frac{1}{4}(\alpha_0-a_0)^2+\frac{1}{2}\sum_{k=1}^{n}[(\alpha_k-a_k)^2+(\beta_k-b_k)^2].$$

显然当右边的正项等于零时，也就是当$\alpha_0=a_0$，$\alpha_k=a_k$，$\beta_k=b_k$ $(k=1,2,\cdots)$时，δ_n^2有最小值。换句话说，用n级三角多项式作函数$f(x)$的近似表达式时，如果这多项式的系数是函数$f(x)$的Fourier 系数，则平方中值误差最小。此时最小平方误差

$$2\varepsilon_n^2=\frac{1}{\pi}\int_{-\pi}^{+\pi}|f(x)|^2dx-\frac{a_0^2}{2}-\sum_{k=1}^{n}(a_k^2+b_k^2).$$

因为$\lim 2\varepsilon_n^2=0(n\to\infty)$，所以有 Parseval 等式(5.40)式成立。因ε_n^2恒正，所以又得到

$$\frac{1}{\pi}\int_{-\pi}^{+\pi}|f(x)|^2dx\geqslant\frac{a_0^2}{2}+\sum_{k=1}^{\infty}(a_k^2+b_k^2),$$

$$\frac{1}{\pi}\int_{-\pi}^{+\pi}|f(x)|^2dx>\frac{a_0^2}{2}+\sum_{k=1}^{n}(a_k^2+b_k^2).$$

B. L_2 模估计

对于有限元解的L_2模估计，我们没有形如(5.31)式的最佳逼近表达式，所以不能直接利用插值函数的误差；观察(5.38)式，线性有限元解的L_2模估计应该是$O(h^2)$阶的。为了得到其最佳的L_2模估计，需利用Nitsche技巧。

设z是辅助问题

$$\begin{cases}Lz=-\dfrac{d}{dx}\left(p\dfrac{dz}{dx}\right)+qz=u-u_h,\\ z(a)=0,\quad z(b)=0\text{或}z'(b)=0\end{cases}\tag{5.41}$$

的解，其相应的弱解方程是

$$a(z,v)=(u-u_h,v),\quad\forall v\in H_{0E}^1,\tag{5.42}$$

其中$a(z,v)$是本章§3.2中所定义的正定有界的双线性泛函。特别地取$v=u-u_h\in H_{0E}^1$，则有

$$\|u-u_h\|_0^2=a(u-u_h,z).\tag{5.43}$$

此外，u_h是有限元解，据(5.30)式有

$$a(u-u_h,v_h)=0,\ \forall v_h\in V_h.\tag{5.44}$$

于是联合(5.43)和(5.44)式，就有

$$\|u-u_h\|_0^2=a(u-u_h,z-v_h).\tag{5.45}$$

利用$a(\cdot,\cdot)$的有界性及H^1模的估计式(5.36)，得

$$\begin{aligned}|a(u-u_h,z-v_h)|&\leqslant M\|u-u_h\|_1\|z-v_h\|_1\\&\leqslant c_1h\|u''\|_0\|z-v_h\|_1,\quad\forall v_h\in V_h,\end{aligned}$$

将此式代入(5.45)式右端，对$v_h \in V_h$取下确界且利用H^1模的估计式(5.36)，则有

$$\|u-u_h\|_0^2 \leqslant c_1 h \|u''\|_0 (\inf_{v_h \in V_h} \|z-v_h\|_1)$$

$$\leqslant c_2 h \|u''\|_0 \|z-z_I\|_1 \leqslant c_3 h^2 \|u''\|_0 \|z''\|_0, \tag{5.46}$$

其中z_I是z的插值函数，c_3是与h无关的常数。

现在来估计$\|z''\|_0$，这可由(5.41)式直接得到。由(5.41)式，有

$$-[pz''+p'z']+qz=u-u_h,$$

$$z''=\frac{1}{p}[-p'z'+qz-(u-u_h)],$$

$$\|z''\|_0 \leqslant \frac{1}{|p_{min}|}[|p'|_{max}\|z'\|_0+|q|_{max}\|z\|_0+\|u-u_h\|_0]$$

$$\leqslant c_4[\|z\|_1+\|u-u_h\|_0]. \tag{5.47}$$

再估计$\|z\|_1$。为此令(5.42)式中的$v=z$，且利用$a(z,z)$的正定性，则有

$$\gamma^2\|z\|_1^2 \leqslant |a(z,z)|=|(u-u_h,z)|.$$

$$\leqslant \|u-u_h\|_0\|z\|_0 \leqslant \|u-u_h\|_0\|z\|_1,$$

从而得到

$$\|z\|_1 \leqslant c_5\|u-u_h\|_0. \tag{5.48}$$

最后，将(5.48)式代入(5.47)式再代入(5.46)式，就得到线性有限元解的L_2模误差估计式：

$$\|u-u_h\|_0 \leqslant ch^2\|u''\|_0.$$

总之，有

定理5.9 若$u \in C^2[a,b]$是(5.33)或(5.34)式的解，u_h是其线性有限元解，则有与h无关的常数c使

$$\|u-u_h\|_0 \leqslant ch^2\|u''\|_0$$

成立。

C. L_∞估计

有限元解的H^1模和L_2模误差估计反映了有限元解在平均的意义下弱收敛于真解u，L_∞最大模估计则反映了u_h的逐点收敛特性，其结果有

定理5.10 若$u\in C^2[a,b]$是问题(5.33)或(5.34)式的解，u_h是其线性有限元解，则有与h无关的常数c使得

$$\begin{cases}|u-u_h|_\infty\leqslant ch^2|\ln h||u''|_\infty,\\|u'-u_h'|_\infty\leqslant ch|\ln h||u''|_\infty.\end{cases}\tag{5.49}$$

证明较难，略去。

4.3 二维椭圆边值问题的线性元误差估计

为简单起见，考虑Poisson方程第一边值问题

$$\begin{cases}-\Delta u=f(x,y),\quad (x,y)\in\Omega,\\u|_{\partial\Omega}=0.\end{cases}$$

假定$f\in C^0(\bar{\Omega})$，$\bar{\Omega}$的边界$\partial\Omega$充分光滑，这足以保证解$u\in C^2(\bar{\Omega})$。为了避免三角剖分Ω_h与Ω的几何误差及数值积分引起的误差，我们进一步假定$\bar{\Omega}$是凸多边形区域，并假定出现在有限元方程中的数值积分都是精确计算的。

设$E_1,E_2,\cdots,E_n$是三角剖分单元；记ρ_i是单元E_i的内切圆直径；h_i是单元E_i的外接圆直径；$h=\max\limits_i h_i$。设u_h是线性有限元解，则据(5.31)式有

$$\|u-u_h\|_1\leqslant\beta\inf\|u-v_h\|_1,\quad \forall v_h\in V_h,$$

其中$V_h\subset H^1(\Omega)$是分片可微整体连续的有限维函数空间。特别地，当选取V_h中的插值函数u_I时，上式仍成立，即有

$$\|u-u_h\|_1\leqslant\beta\|u-u_I\|_1.$$

于是，有限元解H^1模的误差估计归结为插值函数H^1模的误差估计。

为此，先在任一单元E上考虑线性插值函数的误差估计。

设单元E的三个顶点编号为1、2和3，那么在E上，线性元的三个基函数为

$$L_1 = \frac{1}{2S}[(x_2y_3 - x_3y_2) + (y_2 - y_3)x + (x_3 - x_2)y],$$

$$L_2 = \frac{1}{2S}[(x_3y_1 - x_1y_3) + (y_3 - y_1)x + (x_1 - x_3)y], \quad (5.50)$$

$$L_3 = \frac{1}{2S}[(x_1y_2 - x_2y_1) + (y_1 - y_2)x + (x_2 - x_1)y],$$

其中，S表示单元E的面积，$(x_i, y_i)(i = 1,2,3)$是E的顶点坐标；线性有限元函数在E上可表示为

$$u_h^{(E)} = u_1L_1 + u_2L_2 + u_3L_3. \quad (5.51)$$

还设一个线性辅助函数

$$p(x,y) = u(a_1,a_2) + u_x(a_1,a_2)(x - a_1) + u_y(a_1,a_2)(y - a_2), \quad (5.52)$$

其中，(a_1,a_2)是E中的任意一点。因$p(x,y)$是一次多项式函数，故又可用面积坐标表示，即有

$$p(x,y) = p(x_1,y_1)L_1 + p(x_2,y_2)L_2 + p(x_3,y_3)L_3, \quad (5.53)$$

于是有

$$u - u_I = (u - p) + (p - u_I).$$

记D^1表示对x或y的一阶偏导数算子，$D^0 = I$是恒等算子，表示不求偏导数，则

$$|D^su - D^su_I| \leqslant |D^su - D^sp| + |D^sp - D^su_I|$$

$$\leqslant |D^su - D^sp| + \sum_{i=1}^{3}|p(x_i,y_i) - u(x_i,y_i)| \cdot |D^sL_i|, \quad s = 0,1. \quad (5.54)$$

将u在(a_1,a_2)点处作Taylor展开，其展开式的线性函数部分就是(5.52)式中的$p(x,y)$，若记$(b_1,b_2)\in E$是展开式余项中合适的点，则(5.54)式的第一部分可表示为

$$D^0(u-p) = \frac{1}{2}\Big[u_{xx}(b_1,b_2)(x-a_1)^2 + 2u_{xy}(b_1,b_2)(x-a_1)(y-a_2) + u_{yy}(b_1,b_2)(y-a_2)^2\Big],$$

$$D_x^1(u-p) = u_{xx}(b_1,b_2)(x-a_1) + u_{xy}(b_1,b_2)(y-a_2),$$

$$D_y^1(u-p) = u_{xy}(b_1,b_2)(x-a_1) + u_{yy}(b_1,b_2)(y-a_2);$$

再令

$$M_2^{(E)}:=\max(\sup_E|u_{xx}|,\sup_E|u_{xy}|,\sup_E|u_{yy}|),$$

则有

$$\begin{cases}\sup_E|D^0(u-p)|\leqslant 2M_2^{(E)}h_E^2,\\ \sup_E|D_x^1(u-p)|\leqslant 2M_2^{(E)}h_E,\\ \sup_E|D_y^1(u-p)|\leqslant 2M_2^{(E)}h_E.\end{cases}\tag{5.55}$$

对于(5.54)式中的第二部分的有关估计，可由(5.50)式并参见图5.6得到

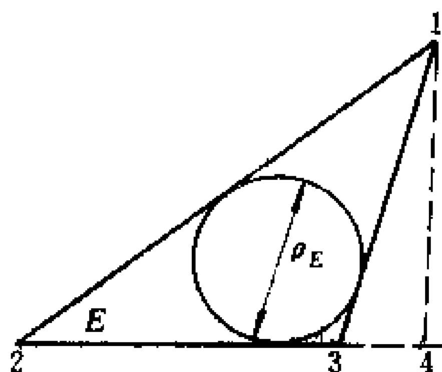


图5.6

$$\begin{cases}\sup_E|D^0L_i|=\sup_E|L_i|\leqslant\sup_E\left|\dfrac{S_i}{S}\right|\leqslant 1,\\ \sup_E|D_x^1L_i|=\sup_E\left|\dfrac{y_2-y_3}{2S}\right|\leqslant\left|\dfrac{|\overrightarrow{23}|}{|\overrightarrow{23}|\cdot|\overrightarrow{14}|}\right|\leqslant\dfrac{1}{\rho_E},\\ \sup_E|D_y^1L_i|=\sup_E\left|\dfrac{x_3-x_2}{2S}\right|\leqslant\left|\dfrac{|\overrightarrow{32}|}{|\overrightarrow{32}|\cdot|\overrightarrow{14}|}\right|\leqslant\dfrac{1}{\rho_E},\\ i=1,2,3.\end{cases}\tag{5.56}$$

将(5.55)和(5.56)式代入(5.54)式，就得到任一单元E上线性插值函数的误差估计式：

$$\sup_E|D^su-D^su_I|\leqslant 2\left[1+3\left(\frac{h_E}{\rho_E}\right)^s\right]M_2^{(E)}h_E^{2-s},\ s=0,1.\tag{5.57}$$

现在对三角剖分加一必要的限制，当剖分无限细分时（即 $h\to 0$），要求对一切单元 E 适合

$$c_1\leqslant\left|\frac{h_E}{\rho_E}\right|\leqslant c_2, \tag{5.58}$$

这相当于所有单元的最小内角大于一正数. 记 $M_2=\max\limits_E\{M_2^{(E)}\}$，那么在区域 Ω 上线性插值函数误差的整体估计可表示为

$$\sup_\Omega|D^s u-D^s u_I|\leqslant cM_2h^{2-s},\quad s=0,1.$$

最后，可得到 Ω 区域上线性插值函数误差的 H^1 模估计

$$\begin{aligned}\|u-u_I\|_1&=\left[\int_\Omega|u-u_I|^2dxdy+\int_\Omega|D_x^1(u-u_I)|^2dxdy\right.\\&\quad\left.+\int_\Omega|D_y^1(u-u_I)|^2dxdy\right]^{1/2}\\&\leqslant\text{meas}(\Omega)^{1/2}\left[\sum_{s=0}^{1}\sup_\Omega|D^s(u-u_I)|\right]\\&\leqslant c\cdot\text{meas}(\Omega)^{1/2}M_2(2+h^2)^{1/2}h,\end{aligned}$$

和线性有限元解的 H^1 模估计：

$$\|u-u_h\|_1\leqslant\beta\|u-u_I\|_1\leqslant C(\Omega)M_2h$$

或
$$\|u-u_h\|_1\leqslant c(\Omega)h\|u\|_2. \tag{5.59}$$

值得指出的是，要得到 H^1 模的误差估计(5.59)，必须满足条件(5.58). 由条件(5.58)式可见，当网格无限细分($h\to 0$)时，若量 h_E/ρ_E 不是一致有界而是无限增大，则有 $\|u-u_h\|_1\nrightarrow 0(h\to 0)$，此时不能保证有限元解收敛；此外，如果剖分中有的单元出现尖角，则 $\max\limits_E\{h_E/\rho_E\}$ 变大，线性有限元解的 H^1 模误差也随之变大，因此在网格剖分时应注意"正规剖分"条件（参见第二章§2），尽量使剖分均匀些.

应用Nitsche技巧，可证明线性有限元解的 L_2 模误差估计. L_∞ 模估计较难，超出本书范围，证明从略，有兴趣的读者可参阅[9～14]和[22]. 知道有限元误差估计的结果对指导数值计算是

有帮助的，因此作为本篇结束，现将有限元解的H^1模、L_2模和L_∞模的误差估计结论列于下面定理中。

定理5.11 设$u\in H^{2,m}(\Omega)$是二阶椭圆边值问题的解，二维凸区域的剖分是正规的，u_h是采用m次元的有限元解，则存在与h无关的常数c，使有下列估计成立：

$$\|u-u_h\|_{s,\Omega}\leqslant ch^{m-s}\|u\|_{m,\Omega},\quad 0\leqslant s\leqslant m-1,$$

$$\|D^s(u-u_h)\|_{\infty,\Omega}\leqslant ch^{m-s-n/2}\|u\|_{m,\Omega},\quad 0\leqslant s< m-n/2;$$

若$u\in H^{2,\infty}(\Omega)$，则有

$$\|D^s(u-u_h)\|_{\infty,\Omega}\leqslant ch^{2-s}|\ln h|\cdot\|u\|_{2,\infty,\Omega},\quad s=0,1.$$

习　题

1．求泛函

$$J(y)=\int_0^1[(y')^2+12xy]dx,\ \ y(0)=0,\ \ y(1)=1$$

在什么曲线上达到极值。

2．求下列泛函在约束条件下取得极值的必要条件：

(1) $$J(y)=\int_a^b F\left(x,y,\frac{dy}{dx},\ \frac{d^2y}{dx^2}\right)dx,$$

$$y(a)=y_a^0,\ \ y(b)=y_b^0,\ \ \frac{dy(a)}{dx}=y_a^1,\ \frac{dy(b)}{dx}=y_b^1;$$

(2) $$J(y)=\int_a^b F\left(x;y_1,\cdots,y_n;\frac{dy_1}{dx},\cdots,\frac{dy_n}{dx}\right)dx,$$

$$y_i(a)=y_{ia}^0,y_i(b)=y_{ib}^0,\ \ i=1,2,\cdots,n.$$

3．在以点$A(x_0,y_0)$和点$B(x_1,y_1)$（设$x_1>x_0,y_0$、$y_1>0$）为端点的所有光滑曲线中，求一曲线使它绕ox轴旋转时所得旋转面的面积最小。

4．在连接定点A、B且与AB线段围成的面积S为定值的光滑曲线中，求出一条曲线使弧段$\overset{\frown}{AB}$最短。

5. 试求出变分问题

$$J(u)=\int_\Omega \frac{1}{2}\left[\left(\frac{\partial u}{\partial x}\right)^2+\left(\frac{\partial u}{\partial y}\right)^2\right]dxdy+\int_{\Gamma_1}\beta u\left(\frac{1}{2}u-t\right)ds\Rightarrow \min,$$

$$u\in M:\ =\{u\in C^2(\Omega)\,|\,u|_{\Gamma_2}=g(x,y),\ \partial\Omega=\Gamma_1\cup\Gamma_2\}$$

（t、β是常数）的解所适合的Euler方程。

6. 有弹性基础的梁，其位能为

$$J(u)=\frac{1}{2}\int_0^l\left[Ku^2+EI\left(\frac{d^2u}{dx^2}\right)^2+2fu\right]dx,$$

其中K、E和I为常数；设边界条件为

$$u(0)=u(l)=0,\quad \frac{du}{dx}(0)=\frac{du}{dx}(l)=0.$$

试利用极小位能原理推导出梁平衡所满足的微分方程定解问题。

7. 试导出下述常微分方程定解问题所对应的变分问题。

$$-\frac{d}{dx}\left(p(x)\frac{du}{dx}\right)+qu=f,\quad a<x<b,$$

其中 $p(x)\in C^1[a,b]$，$p(x)\geqslant \min p(x)>0$，$q(x)$、$f(x)\in C[a,b]$，$q(x)\geqslant 0$。

(1)　$u(a)=\alpha$，$u(b)=\beta$；

(2)　$u(a)=\alpha$，$p(b)\dfrac{du(b)}{dx}+\sigma u(b)=g$；

(3)　$\dfrac{du(a)}{dx}=0$，$p(b)\dfrac{du(b)}{dx}+\sigma u(b)=g$。

8. 利用变分原理证明：常微分方程

$$-\frac{d}{dx}\left[(1+x^2)\frac{du}{dx}\right]+5u=11x^2-3x+2,\quad x\in(0,1)$$

在边界条件$u(0)=u(1)=0$下的解若存在必唯一。

9. 试构造定解问题

$$\begin{cases} -\nabla(k\nabla u)+\dfrac{\partial u}{\partial x}+u^2=f, & (x,y)\in\Omega, \\ u=g, & (x,y)\in\partial\Omega \end{cases}$$

相对应的变分问题，其中

$$\nabla(k\nabla u)=\frac{\partial}{\partial x}\left(k_1\frac{\partial u}{\partial x}\right)+\frac{\partial}{\partial y}\left(k_2\frac{\partial y}{\partial y}\right).$$

10．用虚功原理证明：若定解问题

$$\begin{cases} -\nabla u=f, & (x,y)\in\Omega, \\ \dfrac{\partial u}{\partial n}+\alpha u=0, & (x,y)\in\partial\Omega, \end{cases}$$

$(\alpha>0)$的解存在，则必唯一。

11．在函数类H^1中构造第7题和第9题中所要求的变分问题。

12．设X和B是n阶列向量，A是$n\times n$阶正定对称矩阵，试证明：二次泛函

$$J(X)=\frac{1}{2}(AX,X)-(B,X)$$

在X_0处取得极值的充要条件是

$$AX_0=B.$$

13．用古典Ritz-Galerkin方法求解

$$-\frac{d^2u}{dx^2}=e^x \quad (0<x<1)$$

(1) 对边值条件$u(0)=u(1)=0$，用满足边值条件的试探函数$\varphi(x)=x(1-x)(a+bx)$；

(2) 对边值条件$u(0)=\dfrac{du(1)}{dx}=0$，用满足边值条件的试探函数$\varphi(x)=a(x^2-2x)+b(x^3-3x)$。

14．用线性元手工计算两点边值问题

$$\frac{d^2u}{dx^2}=2,$$
$$u(0)=u(1)=0$$

在[0,1]被2等分和4等分情形下的数值解位移值。

(1) 用外推技巧计算区间中点处的位移近似值，将这中点处的三个近似值与准确解$u=x^2-x$在中点处的值进行比较；

(2) 用超收敛性技巧计算在上述两种分划下，于点1/4、1/2和3/4处的有限元导数值，将这些值与准确解在该点处的导数值进行比较。

15. 证明积分公式

$$\int_0^1 \xi^\alpha(1-\xi)^\beta d\xi=\frac{\alpha!\,\beta!}{(\alpha+\beta+1)!}.$$

16. 用有限元方法数值求解两点边值问题

$$-\frac{d^2u}{dx^2}=f \quad (0<x<1),$$
$$u(0)=a,\ u(1)=b,$$

此时记$x_i(i=1,2,\cdots,N-1)$为内节点，φ_i为节点基函数，$h=1/N$为步长，$a(\cdot,\cdot)$为相应的双线性泛函。

(1) 写出有限元方程组；

(2) 具体计算并写出在单元$[x_i,x_{i+1}]$上的线性元子刚阵

$$\begin{bmatrix} a(\varphi_i,\varphi_i) & a(\varphi_i,\varphi_{i+1}) \\ a(\varphi_{i+1},\varphi_i) & a(\varphi_{i+1},\varphi_{i+1}) \end{bmatrix},$$

这里，总刚阵在一个单元$[x_i,x_{i+1}]$上的表现称为子刚阵；

(3) 具体计算并写出线性元总刚阵中与内节点x_i有关的那一行的3个非零元素；

(4) 利用(2)和(3)的结果，叙述子刚阵叠加形成总刚阵的过程；

(5) 具体计算并写出在单元$[x_i,x_{i+1}]$上的二次元子刚阵

$$\begin{pmatrix} a(\varphi_i,\varphi_i) & a(\varphi_i,\varphi_{i+1/2}) & a(\varphi_i,\varphi_{i+1}) \\ a(\varphi_{i+1/2},\varphi_i) & a(\varphi_{i+1/2},\varphi_{i+1/2}) & a(\varphi_{i+1/2},\varphi_{i+1}) \\ a(\varphi_{i+1},\varphi_i) & a(\varphi_{i+1},\varphi_{i+1/2}) & a(\varphi_{i+1},\varphi_{i+1}) \end{pmatrix},$$

其中，$\varphi_i=(2\xi-1)(\xi-1)$，$\varphi_{i+1/2}=4\xi(1-\xi)$，$\varphi_{i+1}=\xi(2\xi-1)$；

16．具体计算并写出二次元总刚阵中与内节点x_i有关行中的非零元素

$$a(\varphi_{i-1},\varphi_i),a(\varphi_{i-1/2},\varphi_i),a(\varphi_i,\varphi_i),a(\varphi_{i+1/2},\varphi_i),a(\varphi_{i+1},\varphi_i),$$

和单元中点$x_{i+1/2}$有关行中的非零元素

$$a(\varphi_i,\varphi_{i+1/2}),a(\varphi_{i+1/2},\varphi_{i+1/2}),a(\varphi_{i+1},\varphi_{i+1/2}).$$

17．试在[0,1]上构造三次Hermite型插值，其四个基函数$N_0(\xi)$、$N_1(\xi)$、$M_0(\xi)$和$M_1(\xi)$满足下列插值条件：

$$N_0(0)=1,N_0(1)=0,\ \frac{dN_0(0)}{dx}=\frac{dN_0(1)}{dx}=0;$$

$$N_1(1)=1,N_1(0)=0,\frac{dN_1(0)}{dx}=\frac{dN_1(1)}{dx}=0;$$

$$\left.\frac{dM_0(\xi)}{dx}\right|_{\xi=0}=1,\left.\frac{dM_0(\xi)}{dx}\right|_{\xi=1}=0,\ M_0(0)=M_0(1)=0;$$

$$\left.\frac{dM_1(\xi)}{dx}\right|_{\xi=1}=1,\ \left.\frac{dM_1(\xi)}{dx}\right|_{\xi=0}=0,\ M_1(0)=M_1(1)=0.$$

(1) 具体写出单元节点基函数$N_0(\xi)$、$N_1(\xi)$、$M_0(\xi)$和$M_1(\xi)$；

(2) 利用(1)中结果具体写出内节点$x_i\in[a,b]$处的整体节点基函数；

(3) 由(2)中整体节点基函数构成的三次Hermite插值函数在[a,b]上的光滑性如何？

18．设L_1、L_2、L_3为三角形单元e上的面积坐标，e的面积记为S_e，还设p、q、r都是非负整数，试验证

$$\int_e L_1^p L_2^q L_3^r dxdy=\frac{p!\,q!\,r!}{(p+q+r+2)!}2S_e.$$

19. 用直角三角($h_1 \neq h_2$)线性元求解

$$-\Delta u + r(x,y)u = f(x,y), \quad (x,y) \in \Omega$$

的边值问题。

(1) 试计算出以内节点(x_i, y_j)为顶点的某个单元上的子刚阵和子右端项；

(2) 试计算出有限元方程中与内节点(x_i, y_j)有关行中的全部非零元素；

(3) 若采用正三角形线性元，试完成(1)和(2)中的要求。

20. 用直角三角($h_1 = h_2$)线性元解边值问题。

$$-\Delta u = f, \quad (x,y) \in \Omega,$$

$$u = g, \quad (x,y) \in \Gamma_1,$$

$$\frac{\partial u}{\partial n} + u = q, \quad (x,y) \in \Gamma_2,$$

其中$\Gamma_1 \cup \Gamma_2 = \partial\Omega, \bar{\Omega} = [0,a] \times [0,b]$，$\Gamma_2 = [0,a]$。设采用等腰直角三角形分划，试写出有限元方程中与某节点$Q \in \Gamma_2$有关的那一行的全部非零元素。

21. 证明由双二次插值公式和双三次插值公式生成的试探函数空间属于H^1。

22. 编制计算程序，用直角三角线性元求解

$$\begin{cases} -\Delta u = f, & (x,y) \in \Omega = (0,1) \times (0,1), \\ u = g, & (x,y) \in \partial\Omega, \end{cases}$$

此时，将[0,1]均分2^r等份，$r = 1,2,\cdots$，记相应网格步长为hr，相应的数值解为u_{hr}。以定解问题

$$\begin{cases} -\Delta u = \pi^2 \sin\pi x(5\cos 2\pi y - 1), & in\Omega, \\ u|_{\partial\Omega} = 0 \end{cases}$$

为例，其准确解是$u(x,y) = \sin\pi x(\cos 2\pi y - 1)$，具体计算比较下列问题。

(1) 计算u_{h1}、u_{h2}、u_{h3}和u_{h4}，将$\|u_{hr}\|$表示成$C_r(hr)^2$的形式，观察随着网格分划加密，数值解精度在提高的规律，其中

$\|\cdot\|$可分别取为 u_{hr} 按分量绝对值最大的模和Euclid 模，下面(2)和(3)中$\|\cdot\|$的意义也如此；

(2) 用u_{h1}和u_{h2}外推得$\tilde{u}_{h2}$，用u_{h2}和u_{h3}外推得$\tilde{u}_{h3}$，用u_{h3}和和u_{h4}外推得$\tilde{u}_{h4}$，比较$\|u_{hr}\|$和$\|\tilde{u}_{hr}\|$的精度，$r=2,3,4$；

(3) 利用超收敛性技巧，在Ω_{hr}网格上计算得相应的导数近似值u'_{hr}，$r=1$、2、3、4，观察$\|u'_{hr}\|$随着网格细分按$(hr)^2$精度阶增加的规律。

参 考 文 献

[1] D.Gilbarg & N.S. Trudinger著，叶其孝等译，二阶椭圆型偏微分方程，上海科学技术出版社，1981年。

[2] Л.Э. 艾利斯哥尔兹著，李世晋译，变分法，高等教育出版社，1983年。

[3] R.Corandt & D.Hilbert, *Methods of Mathematical Physics*, Vol. 1, 1962年（有中译本）。

[4] 彭旭麟、罗汝梅，变分法及其应用，华中工学院出版社，1984年。

[5] 陈庆益、柳训明，常微分方程及其应用，华中工学院出版社，1984年。

[6] R. A. Adams著，叶其孝等译，索伯列夫空间，人民教育出版社，1983年。

[7] В.И. 斯米尔诺夫著，宋正译，高等数学教程，第二卷第二分册，高等教育出版社，1964年。

[8] P.M. 普伦特著，柴家振等译，样条函数与变分方法，上海科学技术出版社，1980年。

[9] G. Strang& G.Fix, *An analysis of the finite element method*, Prentice-Hall, 1973.

[10] P.G.Ciarlet, *The finite element method for elliptic problems*, Noth-Holland Publishing Company, 1978.

[11] H.R.Schwarz, *Methode der finiten Elemente, Teubner Studienbücher*, Stuttgart, 1980.

[12] 姜礼尚、庞之垣，有限元方法及其理论基础，人民教育出版社，1978年。

[13] 李开泰、黄艾香、黄庆怀，有限元方法及其应用，西安交通大学出版社，1985年。

[14] 陈传淼，有限元方法及其提高精度的分析，湖南科学技术出版社，1982年。

[15] 李荣华、冯果忱，微分方程数值解法，人民教育出版社，1981年。

[16] 蒋叔豪、孙庆新，偏微分方程数值解，浙江大学出版社，1985年。

[17] C.Г. 米赫林著，周先意译，数学物理中的直接方法，高等教育出版社，1959年。

[18] 李庆扬、王能超、易大义，数值分析，华中工学院出版社，1981年。

[19] 易大义、蒋叔豪、李有法，数值方法，浙江科学技术出版社，1984年。

[20] Lin Quen et-al, *Asymptotic expansion for finite element approximations*, IMS-11, 1983.

[21] Lin Quen et-al, *Some expansions of the finite element approximation*, IMS-15, 1984.

[22] 朱起定，有限元超收敛理论，湖南科技出版社，1989年。

[23] 徐长发，关于 *Dirichlet* 边值最小面积问题的有限元解法，华中工学院学报，第九卷，第六期，1981年。

[24] 徐长发，一种求解有限元方程的新方法，华中工学院学报，第13卷第2期，1985年。

[25] 陆金甫，关治，偏微分方程数值解法，清华大学出版社，1987年。

第三篇　解离散微分方程的高效率方法

前两篇介绍了数值求解偏微分方程的两种主要的离散化方法，即差分法和有限元法．这两种方法把求解微分方程定解问题转化为求解离散微分方程的问题．

本篇集中讨论关于高效率高精度求解有限元方程和差分方程的问题．与传统的数值方法相比，一种高效率高精度的数值方法将会几十倍、甚至更有效地节省计算工作量，因此本篇讨论的问题具有明显的理论意义和实用价值．

众所周知，解离散微分方程（代数方程组）的方法一般分为两类，一类是直接解法，另一类是迭代解法．直接解法，如Gauss消元法、列主元消去法、平方根法等等，都是以消元法为基础，用这类方法求解大型稀疏的代数方程组，所需要的存贮量较大，只能解线性问题，而且数值解精度还会受到大量舍入误差的限制．用迭代方法解离散微分方程，则程序设计简单，所需存贮量小，对舍入误差不敏感（这是因为收敛的迭代过程乃是一个不断修正误差的渐近过程），数值解精度较好，且适用于求解非线性离散微分方程；但通常情况下存在收敛速度慢，计算量大的缺点．例如用通常的迭代方法且在正常的精度要求下解二维的椭圆离散方程时，其计算量一般为 $O(N^a)$（$a>1$）量级，其中 N 表示离散方程中的未知量个数，而且这些迭代方法的收敛速度都是与网格分划参数 h 有关的，即网格分划越细，迭代收敛速度越慢．50年代是研究和应用迭代方法的黄金时代，直到60年代，迭代法还是解椭圆离散方程的主要方法；由于舍入误差理论及稀疏矩阵理论方面的一些新成就，直接法在求解离散微分方程方面曾

广泛应用；然而，由于计算量和存贮量都比直接法少以及程序设计简单等原因，迭代法仍旧是解离散微分方程（线性代数方程组，特别是非线性代数方程组）的重要方法。因此，研究和改进现有的迭代法，构造新的更好的迭代法，仍旧是计算数学领域中十分活跃的研究方向之一。人们向往着能构造出某种迭代法，用它求解二维的椭圆离散方程时，其敛速与 h 无关，计算量为 $O(N)$ 量级。这方面的研究与数值代数中的有关研究有所不同，它具有很强的针对性，它可以利用离散的和连续的微分方程的许多固有的性质，它会涉及到网格分划方式，也会涉及到不同网格层次上的数值解之间的关系等等，总之可以利用一切可能利用的信息以达到高效率高精度数值求解离散微分方程的目的。

为了富有启发性且循序渐进地讨论问题，本篇先研究有限元方程与差分方程的某些统一性；然后在此基础上再研究基本迭代法求解椭圆离散微分方程时所存在的种种规律，研究基本迭代法的比较、控制和改进办法，还剖析了基本迭代法产生种种缺点的原因，找到了扬长避短的办法；最后重点讨论各种实用的高效率、高精度求解离散微分方程的迭代方法。

诚然，求解离散抛物型和离散双曲型（隐格式）方程，一般都可归结为求解离散椭圆型方程的求解问题；求解非线性离散微分方程一般可利用线性化方法，也可从求解线性离散微分方程的方法中得到启发；因此，本篇内容主要讨论求解线性椭圆离散方程。另外，虽然目前关于高效率、高精度求解离散微分方程的研究仍在丰富多采地发展着，讨论的问题也很多，但基于本书宗旨和篇幅的原因，本着有利于读者学习、应用和拓广的目的，本篇只介绍几种思想新颖且便于实用的新算法。

第一章　差分格式和有限元格式

本章先简略介绍数值求解椭圆边值问题的差分方法，介绍方便形成和求解差分方程的局部离散差分格式；再介绍在矩形分划下方便形成和求解有限元方程的局部离散有限元格式；还探讨了局部差分格式和有限元格式的某些统一性。

§1　解 Poisson 方程的差分方法

考虑解 Poisson 方程

$$Lu = -\Delta u = f, \qquad (x, y) \in \Omega \tag{1.1}$$

的差分方法，其中定解区域 $\bar{\Omega}$ 是有界凸平面区域，$\bar{\Omega} = \Omega \cup \partial\Omega$，区域边界 $\partial\Omega$ 是由分段光滑曲线组成。椭圆型偏微分方程(1.1)的**定解条件**有三类：

第一类边值条件　$u|_{\partial\Omega} = \alpha(x, y)$；　(1.2a)

第二类边值条件　$\left.\dfrac{\partial u}{\partial n}\right|_{\partial\Omega} = \beta(x, y)$；　(1.2b)

第三类边值条件　$\left[\dfrac{\partial u}{\partial n} + k(x, y)u\right]_{\partial\Omega} = \gamma(x, y)$；　(1.2c)

其中 α、β 和 γ 都是已知的连续函数，$k \geq 0$，$\partial u/\partial n$ 是 $\partial\Omega$ 的外法向导数。

现取步长为 h_1 和 h_2 的两族平行直线

$$x = x_i = ih_1, \quad i = 0, \pm 1, \pm 2, \cdots,$$
$$y = y_j = jh_2, \quad j = 0, \pm 1, \pm 2, \cdots$$

来剖分区域 $\bar{\Omega}$，这就是大家熟知的**矩形网格剖分**。这两族平行直线的交点 (x_i, y_j) 称为节点。当 $(x_i, y_j) \in \Omega$ 时，称为**内**

节点（或**内点**）；如果某内节点的上下左右 4 个相邻节点都属于 Ω，则称其为**正则内点**，否则称为**非正则内点**；若 $(x_i, y_j)\in\partial\Omega$，则称其为**界点**。图 1.1 用"○"号标注的点是正则内点，用"×"号标注的点是非正则内点，用"•"号标注的点是界点。记 $\Omega_h=\{(x_i, y_j)\mid(x_i, y_j)\in\Omega\}$，$\partial\Omega_h=\{$界点全体$\}$，则 $\bar{\Omega}_h=\Omega_h\cup\partial\Omega_h$。

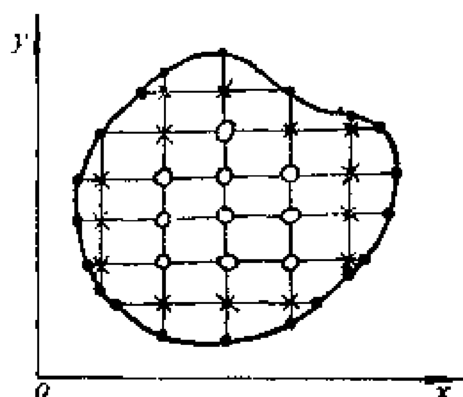


图 1.1

在差分方法中，人们常用 $u(x_i, y_j)$ 表示 $u(x, y)$ 在节点 (i, j) 处的值；并用 u_h、f_h 表示相应的**网格函数**且在 $(i, j)\in\Omega_h$ 处 $u(x_i, y_j)\approx u_h(x_i, y_j)=u_{ij}$，$f(x_i, y_j)=f_h(x_i, y_j)=f_{ij}$。

1.1 五点差分格式

设 (i, j) 是正则内点，解函数 $u(x, y)$ 足够光滑，由 Taylor 展式则有

$$\frac{1}{h_1^2}[u(x_{i+1}, y_j)-2u(x_i, y_j)+u(x_{i-1}, y_j)]$$

$$=\frac{\partial^2 u(x_i, y_j)}{\partial x^2}+\frac{h_1^2}{12}\frac{\partial^4 u(x_i, y_j)}{\partial x^4}+\frac{h_1^4}{360}\frac{\partial^6 u(x_i, y_j)}{\partial x^6}+O(h_1^6), \tag{1.3}$$

$$\frac{1}{h_2^2}[u(x_i, y_{j+1})-2u(x_i, y_j)+u(x_i, y_{j-1})]$$

$$=\frac{\partial^2 u(x_i, y_j)}{\partial y^2}+\frac{h_2^2}{12}\frac{\partial^4 u(x_i, y_j)}{\partial y^4}+\frac{h_2^4}{360}\frac{\partial^6 u(x_i, y_j)}{\partial y^6}+O(h_2^6). \tag{1.4}$$

用（1.3）、（1.4）式中的沿 x 和 y 方向的二阶中心差商直接代替方程（1.1）中的 u_{xx} 和 u_{yy}，就得到**差分方程** $L_h u_h = -\Delta_h u_h = f_h$，即

$$-\frac{1}{h_1^2}[u_{i+1,j}-2u_{ij}+u_{i-1,j}]$$

$$-\frac{1}{h_2^2}[u_{i,j+1}-2u_{ij}+u_{i,j-1}]=f_{ij}. \tag{1.5}$$

由于（1.5）式中只出现 u 在 (i, j) 及其 4 个邻点处的值，故称此差分方程为**5点差分格式**，参见图1.2。显然，由(1.3)和（1.4）式可知 5 点差分格式的**截断误差**

$$R_{ij}=\Delta u(x_i, y_j)-\Delta_h u(x_i, y_j)$$

$$=\frac{-1}{12}\left[h_1^2\frac{\partial^4 u}{\partial x^4}+h_2^2\frac{\partial^4 u}{\partial y^4}\right]_{(i,j)}+O(h^4)=O(h^2),$$

其中 $h=\max\{h_1, h_2\}$。截断误差反映了差分算子对微分算子的相容逼近，也反映了差分解的局部误差状况。

特别地取正方形网格（$h_1=h_2=h$），差分方程（1.5）则简化为

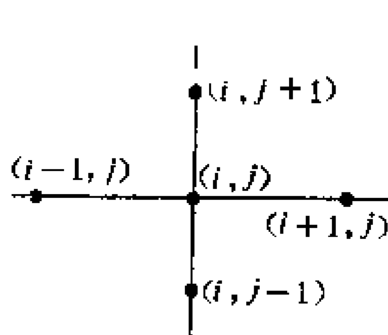


图 1.2

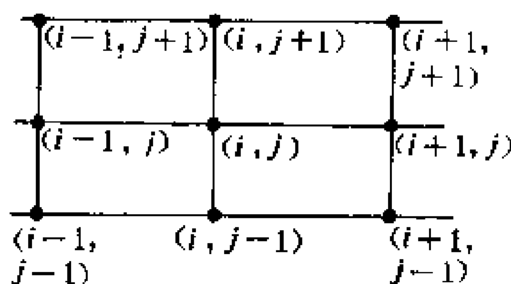


图 1.3

$$u_{ij}-\frac{1}{4}(u_{i-1,j}+u_{i+1,j}+u_{i,j-1}+u_{i,j+1})=\frac{h^2}{4}f_{ij}. \tag{1.6}$$

例如，用 5 点差分格式数值求解模型问题

$$\begin{cases}-\Delta u=f, & (x,\ y)\in\Omega=(0,\ 1)\times(0,\ 1),\\ u=g, & (x,\ y)\in\partial\Omega,\end{cases}$$

可先将 $[0,\ 1]\times[0,\ 1]$ 等分 $2^N\times 2^N$ 等分，步长 $h=1/2^N$；节点编号按从左到右从下向上的顺序即**坐标式编号**；采用 5 点差分格式 (1.6)，就得到差分方程组，其系数矩阵是 5 对角的；解这个差分方程组，即可求得节点 $(i,\ j)$ 处的差分解 $u_{ij}(i,\ j=1,\ 2,\ \cdots,\ 2^N+1)$。

1.2 9点差分格式

为了提高差分算子逼近的阶，仍可用 Taylor 展式。将 (1.3)和 (1.4) 式相加，就有

$$\begin{aligned}
&\Delta_h u(x_i,\ y_j)\\
&=\Delta u(x_i,\ y_j)+\frac{1}{12}\left[h_1^2\frac{\partial^4 u(x_i,\ y_j)}{\partial x^4}+h_2^2\frac{\partial^4 u(x_i,\ y_j)}{\partial y^4}\right]\\
&\quad+O(h^4)\\
&\qquad=\Delta u(x_i,\ y_j)+\frac{1}{12}\left[h_1^2\frac{\partial^2}{\partial x^2}+h_2^2\frac{\partial^2}{\partial y^2}\right]\\
&\qquad\quad\times\left[\frac{\partial^2 u(x_i,\ y_j)}{\partial x^2}+\frac{\partial^2 u(x_i,\ y_j)}{\partial y^2}\right]\\
&\qquad\quad-\frac{h_1^2+h_2^2}{12}\ \frac{\partial^4 u(x_i,\ y_j)}{\partial x^2\partial y^2}+O(h^4)\\
&=-f(x_i,\ y_j)-\frac{1}{12}\left[h_1^2\frac{\partial^2 f(x_i,y_j)}{\partial x^2}+h_2^2\frac{\partial^2 f(x_i,\ y_j)}{\partial y^2}\right]\\
&\quad-\frac{h_1^2+h_2^2}{12}\ \frac{\partial^4 u(x_i,\ y_j)}{\partial x^2\partial y^2}+O(h^4).
\end{aligned}$$

又

$$\frac{\partial^4 u(x_i, y_j)}{\partial x^2 \partial y^2}$$

$$= \frac{u_{xx}(x_i, y_{j+1}) - 2u_{xx}(x_i, y_j) + u_{xx}(x_i, y_{j-1})}{h_2^2} + O(h_2^2)$$

$$= \frac{1}{h_1^2 h_2^2}\Big[u(x_{i+1}, y_{j+1}) - 2u(x_i, y_{j+1}) + u(x_{i-1}, y_{j+1})$$

$$- 2(u(x_{i+1}, y_j) - 2u(x_i, y_j) + u(x_{i-1}, y_j))$$

$$+ u(x_{i+1}, y_{j-1}) - 2u(x_i, y_{j-1}) + u(x_{i-1}, y_{j-1})\Big] + O(h^2).$$

因此

$$\Delta_h u(x_i, y_j) + \frac{1}{12}[4u(x_i, y_j) - 2u(x_{i-1}, y_j)$$

$$- 2u(x_{i+1}, y_j) - 2u(x_i, y_{j-1}) - 2u(x_i, y_{j+1})$$

$$+ u(x_{i-1}, y_{j-1}) + u(x_{i+1}, y_{j-1}) + u(x_{i+1}, y_{j+1})$$

$$+ u(x_{i-1}, y_{j+1})](h_1^2 + h_2^2)/h_1^2 h_2^2 = -f(x_i, y_j)$$

$$- \frac{1}{12}\left[h_1^2 \frac{\partial^2 f(x_i, y_j)}{\partial x^2} + h_2^2 \frac{\partial^2 f(x_i, y_j)}{\partial y^2}\right] + O(h^4).$$

舍去截断误差项 $R_{ij} = O(h^4)$，便得到近似 Poisson 方程的 **9点差分格式**：

$$-\frac{1}{12}(h_1^2 + h_2^2)u_{i-1,j+1} + \left(\frac{1}{6}h_2^2 - \frac{5}{6}h_1^2\right)u_{i,j+1}$$

$$-\frac{1}{12}(h_1^2 + h_2^2)u_{i+1,j+1} + \left(\frac{1}{6}h_1^2 - \frac{5}{6}h_2^2\right)u_{i-1,j}$$

$$+\frac{5}{3}(h_1^2 + h_2^2)u_{ij} + \left(\frac{1}{6}h_1^2 - \frac{5}{6}h_2^2\right)u_{i+1,j}$$

$$-\frac{1}{12}(h_1^2 + h_2^2)u_{i-1,j-1} + \left(\frac{1}{6}h_2^2 - \frac{5}{6}h_1^2\right)u_{i,j-1}$$

$$-\frac{1}{12}(h_1^2 + h_2^2)u_{i+1,j-1}$$

$$=h_1^2h_2^2\left[f_{ij}+\frac{1}{12}h_1^2f_{xx}(x_i, y_j)+\frac{1}{12}h_2^2f_{yy}(x_i, y_j)\right].$$

(1.7)

当 $h_1=h_2$ 时，相应有

$$-u_{i-1,j+1}-4u_{i,j+1}-u_{i+1,j+1}-4u_{i-1,j}+20u_{ij}$$
$$-4u_{i+1,j}-u_{i-1,j-1}-4u_{i,j-1}-u_{i+1,j-1}$$
$$=6h^2\left[f_{ij}+\frac{1}{12}h^2(f_{xx}(x_i, y_j)+f_{yy}(x_i, y_j))\right].$$

(1.8)

构造9点差分格式所涉及的邻节点关系见图1.3.

1.3 特殊9点差分格式

同样可以在 $(i+1/2, j)$ 和 $(i-1/2, j)$ 处构造 x 方向的二阶中心差商，再取其平均值代替 u_{xx}，在 $(i, , j+1/2)$ 和 $(i, j-1/2)$ 处构造 y 方向的二阶中心差商，取其平均值代替 u_{yy}，则得到**特殊9点格式**：

$$-\left[\frac{u_{i+1,j}-2u_{i+1/2,j}+u_{ij}}{2(h_1/2)^2}+\frac{u_{i-1,j}-2u_{i-1/2,j}+u_{ij}}{2(h_1/2)^2}\right]$$
$$\left[\frac{u_{i,j+1}-2u_{i,j+1/2}+u_{ij}}{2(h_2/2)^2}+\frac{u_{i,j-1}-2u_{i,j-1/2}+u_{ij}}{2(h_2/2)^2}\right]$$
$$=f_{ij}.\qquad(1.9)$$

易知，特殊 9 点格式的截断误差为 $O((h/2)^2)$.

1.4 一般椭圆型方程的差分格式

设线性椭圆型方程具有一般形式：

$$Lu=L\left(x, y, \frac{\partial u}{\partial x}, \frac{\partial u}{\partial y}, \frac{\partial^2 u}{\partial x^2}, \frac{\partial^2 u}{\partial y^2}\right)u(x, y)$$
$$=f(x, y),$$

若解函数足够光滑，则在正则内点（i，j）处可利用 Taylor 展开式构造一阶向前、向后、中心差商和二阶中心差商：

$$\left(\frac{\delta_x^<}{h_1}\right)u_{ij}=\frac{u_{i+1,j}-u_{ij}}{h_1}=\frac{\partial u(x_i,\ y_j)}{\partial x}+O(h_1),$$

$$\left(\frac{\delta_x^>}{h_1}\right)u_{ij}=\frac{u_{i-1,j}-u_{ij}}{-h_1}=\frac{\partial u(x_i,\ y_j)}{\partial x}+O(h_1),$$

$$\left(\frac{\delta_x^=}{2h_1}\right)u_{ij}=\frac{u_{i+1,j}-u_{i-1,j}}{2h_1}=\frac{\partial u(x_i,\ y_j)}{\partial x}+O(h_1^2),$$

$$\left(\frac{\delta_y^<}{h_2}\right)u_{ij}=\frac{u_{i,j+1}-u_{ij}}{h_2}=\frac{\partial u(x_i,\ y_j)}{\partial y}+O(h_2),$$

$$\left(\frac{\delta_y^>}{h_2}\right)u_{ij}=\frac{u_{i,j-1}-u_{ij}}{-h_2}=\frac{\partial u(x_i,\ y_j)}{\partial y}+O(h_2),$$

$$\left(\frac{\delta_y^=}{2h_2}\right)u_{ij}=\frac{u_{i,j+1}-u_{i,j-1}}{2h_2}=\frac{\partial u(x_i,\ y_j)}{\partial y}+O(h_2^2),$$

$$\left(\frac{\delta_x^2}{h_1^2}\right)u_{ij}=\frac{u_{i-1,j}-2u_{ij}+u_{i+1,j}}{h_1^2}=\frac{\partial^2 u(x_i,\ y_j)}{\partial x^2}+O(h_1^2),$$

$$\left(\frac{\delta_y^2}{h_2^2}\right)u_{ij}=\frac{u_{i,j-1}-2u_{ij}+u_{i,j+1}}{h_2^2}=\frac{\partial^2 u(x_i,\ y_j)}{\partial y^2}+O(h_2^2).$$

用差商或其组合直接替换微分方程中相应阶数的微商，便可得到在正则内点处近似微分方程的差分方程：

$$L_h u_h=L_h\left(ih_1,\ jh_2,\ \frac{\delta_x^=}{2h_1},\ \frac{\delta_y^=}{2h_2},\ \frac{\delta_x^2}{h_1^2},\ \frac{\delta_y^2}{h_2^2}\right)u_{ij}=f_{ij}.$$

因此，在正则内点处直接构造差分方程无论如何是一件不难的工作，差分方法中的难点和技巧性强的地方在于估计差分解的误差、边值条件的处理、提高差分解精度及差分方程的高效率求解等问题。本书中不讲述有关差分解的误差估计问题，有兴趣的读者请参阅[1]。

1.5 边值条件的处理

一、定解区域是矩形的情形

1. 第一类边值条件的处理

第一类边值条件很容易处理，只要取 $u_{ij}=\alpha_{ij}$，$(x_i, y_j)\in\partial\Omega_h$ 就行了。

2. 第二、三类边值条件的处理

设矩形区域已被剖分为 $N\times M$ 等分，$1\leqslant i\leqslant N+1$，$1\leqslant j\leqslant M+1$。为了处理第二、三类边值条件（1.2$b$）和（1.2$c$），可在 $\bar{\Omega}_h$ 四周再增加一排节点。例如在界点上用一阶中心差商离散（1.2b）式，就有

$$\begin{cases}\left(\dfrac{\partial u}{\partial n}\right)_{1,j}\approx\dfrac{u_{2,j}-u_{0,j}}{2h_1}=\beta_{1,j}, & 1\leqslant j\leqslant M+1,\\ \left(\dfrac{\partial u}{\partial n}\right)_{N+1,j}\approx\dfrac{u_{N+2,j}-u_{N,j}}{2h_1}=\beta_{N+1,j}, & \\ \left(\dfrac{\partial u}{\partial n}\right)_{i,1}\approx\dfrac{u_{i,0}-u_{i,2}}{2h_2}=\beta_{i,1}, & 1\leqslant i\leqslant N+1,\\ \left(\dfrac{\partial u}{\partial n}\right)_{i,M+1}\approx\dfrac{u_{i,M+2}-u_{i,M}}{2h_2}=\beta_{i,M+1}. & \end{cases}\tag{1.10}$$

在上面诸式中，出现了一些附加的节点函数值，我们可以用内节点处的5点差分格式和（1.10）式消去这些附加的节点函数值。例如令（1.5）式中的 $i=1$，就有

$$-\left[\frac{u_{2j}-2u_{1j}+u_{0j}}{h_1^2}+\frac{u_{1,j+1}-2u_{1j}+u_{1,j-1}}{h_2^2}\right]=f_{1j},$$

再与（1.10）式中的第一式联立就能消去 u_{0j}，得

$$-\left[\frac{2u_{2j}-2u_{1j}+2h_1\beta_{1j}}{h_1^2}+\frac{u_{1,j+1}-2u_{1j}+u_{1,j-1}}{h_2^2}\right]=f_{1j}.$$

这样就可以补全第二边界点处的差分方程了。

第三边值条件的处理方法与第二边值条件的处理办法类似。

二、一般定解区域情形

对于不规则的定解区域，各类边值条件的处理要比矩形区域情形复杂一些，有些还缺乏有效的方法。

1．第一类边值条件的处理

非正则内点处也需要有合适的差分格式，它涉及到界点数值的处理问题。如果界点正好是网格节点，如图 1.4 中的 S 点，那

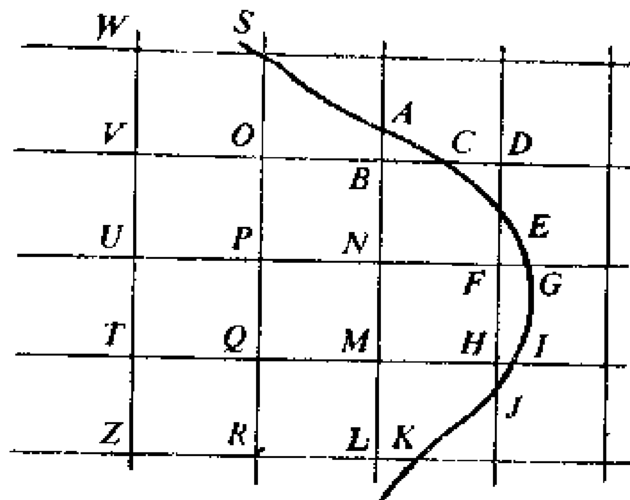


图 1.4

么可将 S 点处的边界函数值直接代入非正则内点处的差分格式中；如果界点不是网格节点，可用不同的办法列写非正则内点处的差分格式，常用的处理办法有下面三种。

（1）修改边界法

对非正则内点，其差分格式一律采用 5 点格式型，这就需要判断每个非正则内点与其一切邻界点的轴向距离关系。如果界点与非正则内点的 y 轴向距离 $\leqslant h_2/2$，就将此边界点沿 y 轴就近迁移到该非正则内点处，也就是说，直接将此非正则内点作为边界点；否则就将 y 轴上处于界外的一个就近的网格节点作为边界点；这两种情况都用边界函数在变动后的界点处的函数值作为边界节点值。在 x 轴向修改边界点的办法同样如此。处理结果见图1.4，即将原来的部分边界 SACEGIJK 修改为 SBDFHL，

即用斜线段和轴向线段作为修改后的边界；对边界值近似的精度只有 $O(h)$ 阶；但这种非正则内点差分格式仍保持了差分方程组的对称性。随着网格不断细密化，修改后的边界就不断地逼近真实边界了。

（2）线性插值法

非正则内点处的差分格式由该点左右两个邻点或上下两个邻点或上下左右4个邻点函数值作线性插值的办法来构造。例如图1.4中，B点是非正则内点，设 $AB=\bar{h}_2$，$BC=\bar{h}_1$；那么

x方向作线性插值，有 $u_B=\dfrac{h_1u_C+\bar{h}_1u_0}{h_1+\bar{h}_1}$；

y方向作线性插值，有 $u_B=\dfrac{h_2u_A+\bar{h}_2u_N}{h_2+\bar{h}_2}$；

x和y方向作线性插值，有

$$u_B=\frac{1}{2}\frac{h_1u_C+\bar{h}_1u_0}{h_1+\bar{h}_1}+\frac{1}{2}\frac{h_2u_A+\bar{h}_2u_N}{h_2+\bar{h}_2}.$$

用线性插值的办法所构造的非正则内点差分格式可具有 $O(h^2)$ 阶的截断误差，但非正则内点处的差分格式与正则内点处的差分格式在形式上有区别，从而它破坏了最后要求解的线性差分方程组的系数矩阵的对称性。

（3）不等距5点格式

仍以构造非正则内点B处（见图1.4）的差分格式为例，因为

$$\left[\frac{\partial^2u}{\partial x^2}\right]_B=\frac{1}{2}(h_1+\bar{h}_1)\left(\frac{u_C-u_B}{\bar{h}_1}-\frac{u_B-u_O}{h_1}\right)+O(h_1),$$

$$\left[\frac{\partial^2u}{\partial y^2}\right]_B=\frac{1}{2}(h_2+\bar{h}_2)\left(\frac{u_A-u_B}{\bar{h}_2}-\frac{u_B-u_N}{h_2}\right)+O(h_2),$$

这样可得到一个截断误差为 $O(h)$ 的不等距5点格式：

$$-\frac{1}{2}(h_1+\bar{h}_1)\left(\frac{u_C-u_B}{\bar{h}_1}-\frac{u_B-u_O}{h_1}\right)$$

$$-\frac{1}{2}(h_2+\bar{h}_2)\left(\frac{u_A-u_B}{\bar{h}_2}-\frac{u_B-u_N}{h_2}\right)=f_B.$$

这样处理在实际计算是可行的，但还是会破坏代数方程组系数矩阵的对称性。若想保持这种对称性，可将上式作适当修改，即令$\bar{h}_1=h_1$，$\bar{h}_2=h_2$，上式就简化为

$$-\frac{1}{h_1^2}(u_C-2u_B+u_O)-\frac{1}{h_2^2}(u_A-2u_B+u_N)=f_B.$$

此时，截断误差阶降为$O(1)$，尽管如此，这在实际计算中也是可行的。

值得指出的是，虽然非正则内点差分格式有时破坏了差分方程组系数阵的对称性，但这并不影响用迭代法求解差分方程组的有效性和存贮量；若用直接法求解差分方程组，这种不对称性则会严重影响计算量和存贮量。

2。第二、三类边值条件的处理

(1) 边界点是节点的情形

如图1.5所示，P是界点又是网格节点，外法线为n，此时P点处的第三类边值条件(1.2c)可离散化为

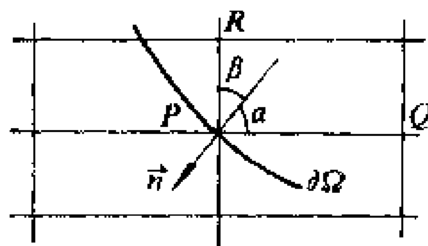


图 1.5

$$\frac{1}{h_1}(u_P-u_Q)\cos\alpha+\frac{1}{h_2}(u_P-u_R)\cos\beta+k_Pu_P=\gamma_P;$$

当外法线指向x轴负向或指向y轴负向时，则上式相应简化为

$$\frac{1}{h_1}(u_P - u_Q) + k_P u_P = \gamma_P,$$

或

$$\frac{1}{h_2}(u_P - u_Q) + k_P u_P = \gamma_P.$$

(2) 边界不经过网格节点的情形

如图 1.6 所示，边界点M不是网格节点，非正则内点P处可使用不等距 5 点格式，边值处理时需补上第二、三类边界点M处的差分格式.

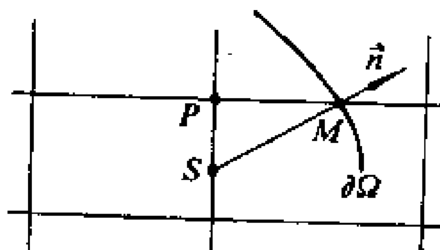


图 1.6

根据 (1.2c) 式在M点的表现，容易写出M点处的离散格式：

$$\frac{1}{|MS|}(u_M - u_S) + k_M u_M = \gamma_M,$$

其中 u_S 可利用内插方法获值.

值得指出的是，在处理第二、三边值条件时，一定要注意外法线矢量的方向，否则会造成错误. 另外，**积分差分法**也是处理第二、三边值条件的有效方法，有兴趣的读者请参阅[1].

§2 差分格式与有限元格式的某些统一性

2.1 差分、有限元局部离散格式

只要观察一下差分方程组中与某个网格节点对应的那一行，就不难发现该行中所涉及到的网格节点函数值，都是 (i, j) 节点及其邻节点处的网格函数值. 所以差分格式是一种**局部离散格**

式；若每个网格节点处都使用某种差分局部离散格式，则所有节点处的局部离散格式的组合便是差分方程组。

为了统一表示这种局部离散格式，设 $x=(ih_1,\ jh_2)\in\Omega_h$，$\mathbf{k}=(\mathbf{k}_1,\ \mathbf{k}_2)$，$h=(h_1,\ h_2)$，$\mathbf{k}\in V$，例如 $V=\{(-1,\ 1),\ (0,\ 1),\ (1,\ 1),\ (-1,\ 0),(0,\ 0),\ (1,\ 0),(-1,\ -1),(0,-1),\ (1,\ -1)\}$，$V$ 称为局部离散格式的**指标集**，它是一个有限集。这样，与 x 相邻的节点可统一写成 $x+\mathbf{k}h$ 的形式，某节点 x 处的差分方程或局部差分格式可表示为

$$[S_{\mathbf{k}}]u_h(x)=\sum_{\mathbf{k}\in V}s_{\mathbf{k}}u_h(x+\mathbf{k}h)=F_{\mathbf{k}}(x),\quad x\in\Omega_h, \tag{1.11}$$

其中：

$$\begin{aligned}&\sum_{\mathbf{k}\in V}s_{\mathbf{k}}u_h(x+\mathbf{k}h)\\&=s_{-1,1}u_{i-1,j+1}+s_{0,1}u_{i,j+1}+s_{1,1}u_{i+1,j+1}\\&\quad+s_{-1,0}u_{i-1,j}+s_{0,0}u_{ij}+s_{1,0}u_{i+1,j}\\&\quad+s_{-1,-1}u_{i-1,j-1}+s_{0,-1}u_{i,j-1}+s_{1,-1}u_{i+1,j-1},\end{aligned} \tag{1.12}$$

$$[S_{\mathbf{k}}]=\begin{pmatrix}s_{-1,1}&s_{0,1}&s_{1,1}\\s_{-1,0}&s_{0,0}&s_{1,0}\\s_{-1,-1}&s_{0,-1}&s_{1,-1}\end{pmatrix}, \tag{1.13}$$

$F_{\mathbf{k}}(x)$ 相应于离散格式的具体形式而定。

按（1.11）~（1.13）式定义的局部离散格式表明了在矩形网格剖分下（节点按坐标形式编号），某节点（编号为$(i,\ j)$）处的差分方程所涉及到的那些网格节点的分布情况，以及这些节点离散值在差分方程中的加权规律。读者以后将会看到，用局部离散格式统一描述**差分格式**和**有限元格式**是方便的，用局部离散格式也方便于差分方程组和有限元方程组的形成及其迭代求解，用局部离散格式还可方便地用**局部 Fourier 分析方法**分析差分格式和有限元格式的收敛速度。

下面将 Poisson 方程的各种差分格式表示为局部离散格式。由（1.5）式定义的 **5点差分格式** 可用局部离散格式表示为

$$
\begin{cases}
[S_k]_{5h}u_h(x) = F_{k5h}(x), \quad \forall x \in \Omega_h, \\
[S_k]_{5h} = \begin{pmatrix} 0 & -h_1^2 & 0 \\ -h_2^2 & 2(h_1^2+h_2^2) & -h_2^2 \\ 0 & -h_1^2 & 0 \end{pmatrix}, \\
F_{k5h}(x) = h_1^2 h_2^2 f(x);
\end{cases}
\tag{1.14}
$$

对于正方形分划情形的5点差分格式，可表示为

$$
\begin{cases}
[S_k]_5 u_h(x) = F_{k5}(x), \quad \forall x \in \Omega_h, \\
[S_k]_5 = \begin{pmatrix} 0 & -1 & 0 \\ -1 & 4 & -1 \\ 0 & -1 & 0 \end{pmatrix}, \quad F_{k5}(x) = h^2 f(x).
\end{cases}
\tag{1.15}
$$

同样，对矩形分划和正方形分划的形如（1.7）和（1.8）式的**9点差分格式**，可用局部离散格式表示为

$$
\begin{cases}
[S_k]_{9h}u_h(x) = F_{k9h}(x), \quad \forall x \in \Omega_h, \\
[S_k]_{9h} = \begin{pmatrix} \frac{-1}{12}(h_1^2+h_2^2) & \frac{1}{6}h_2^2 - \frac{5}{6}h_1^2 & \frac{-1}{12}(h_1^2+h_2^2) \\ \frac{1}{6}h_1^2 - \frac{5}{6}h_2^2 & \frac{5}{3}(h_1^2+h_2^2) & \frac{1}{6}h_1^2 - \frac{5}{6}h_2^2 \\ \frac{-1}{12}(h_1^2+h_2^2) & -\frac{1}{6}h_2^2 - \frac{5}{6}h_1^2 & \frac{-1}{12}(h_1^2+h_2^2) \end{pmatrix}, \\
F_{k9h}(x) = h_1^2 h_2^2\left[f(x) + \frac{1}{12}h_1^2 f''_{x_1}(x) + \frac{1}{12}h_2^2 f''_{x_2}(x)\right] \\
\qquad \approx h_1^2 h_2^2\left(1 - \frac{1}{12}[S_k]_5\right) f(x);
\end{cases}
\tag{1.16}
$$

和

$$
\begin{cases}
[S_k]_9 u_h(x) = F_{k9}(x), \quad \forall x \in \Omega_h, \\
[S_k]_9 = \frac{1}{6}\begin{pmatrix} -1 & -4 & -1 \\ -4 & 20 & -4 \\ -1 & -4 & -1 \end{pmatrix}, \\
F_{k9} = h^2\left(1 - \frac{1}{12}[S_k]_5\right) f(x).
\end{cases}
\tag{1.17}
$$

特殊9点差分格式（1.9）表示为局部离散格式的情形是

$$\begin{cases}[S_k]_{9h*}u_h(x)=F_{k9h*}(x), \quad \forall x\in\Omega_h,\\ [S_k]_{9h*}=2\begin{bmatrix} & & -h_1^2 & & \\ & & 2h_1^2 & & \\ -h_2^2 & 2h_2^2 & -2(h_1^2+h_2^2) & 2h_2^2 & -h_2^2\\ & & 2h_1^2 & & \\ & & -h_1^2 & & \end{bmatrix},\\ F_{k9h*}(x)=h_1^2h_2^2f(x);\end{cases} \tag{1.18}$$

当 $h_1=h_2=h$ 时，有

$$\begin{cases}[S_k]_{9*}u_h(x)=F_{k9*}(x), \quad \forall x\in\Omega_h,\\ [S_k]_{9*}=2\begin{bmatrix} & & -1 & & \\ & & 2 & & \\ -1 & 2 & -4 & 2 & -1\\ & & 2 & & \\ & & -1 & & \end{bmatrix},\\ F_{k9*}(x)=h^2f(x).\end{cases} \tag{1.19}$$

下面进一步讨论在矩形网格分划下解 Poisson 方程的种种**有限元局部离散格式**.

由第二篇可知，线性椭圆边值问题的有限元方程组是一个如下形式的线性代数方程组

$$\begin{pmatrix} a(\varphi_1, \varphi_1) & \cdots & a(\varphi_n, \varphi_1)\\ \cdots & \cdots & \cdots\\ a(\varphi_1, \varphi_n) & \cdots & a(\varphi_n, \varphi_n)\end{pmatrix}\begin{pmatrix}u_1\\ \cdots\\ u_n\end{pmatrix}=\begin{pmatrix}(f, \varphi_1)\\ \cdots\\ (f, \varphi_n)\end{pmatrix},$$

其中 φ_i 是整体节点基函数，网格节点是按**整体累加编号规则**编号的. 这个方程组的系数矩阵，即总刚阵，是一个大型稀疏矩阵，其中的任意一行只有少数几个非零元素. 总刚阵的这种稀疏特性是由节点基函数的局部非零特性决定的. 总刚阵某行中的零元素与非零元素的分布规律是

$$a(\varphi_i, \varphi_j)\begin{cases} =0, & \text{当 } j \text{ 不是 } i \text{ 及其邻节点时,} \\ \neq 0, & \text{当 } j \text{ 是 } i \text{ 及其邻节点时.} \end{cases}$$

因此和差分格式一样，有限元格式也是一种局部的离散格式。在任意一个内节点处使用这种离散格式，就得到有限元方程组中有关该节点的方程，有限元方程组则是对所有内节点施行某种离散格式的结果。

为了描述三角线性元局部离散格式，不妨假设采用强正规三角分划；设单元三个角度参数分别为 α_1、α_2 和 α_3；记 $c_i = \mathrm{ctg}\alpha_i (i=1, 2, 3)$，$c = c_1 + c_2 + c_3$，$S$ 是单元面积；并不失一般性地假设每个内节点有 6 个邻点（特别当采用网格分半加密办法时，更是如此），参见图1.7。利用简单的仿射变换即可将图 1.7 变为图 1.8，所以只要在如图 1.7 的分划中采用斜交网格且

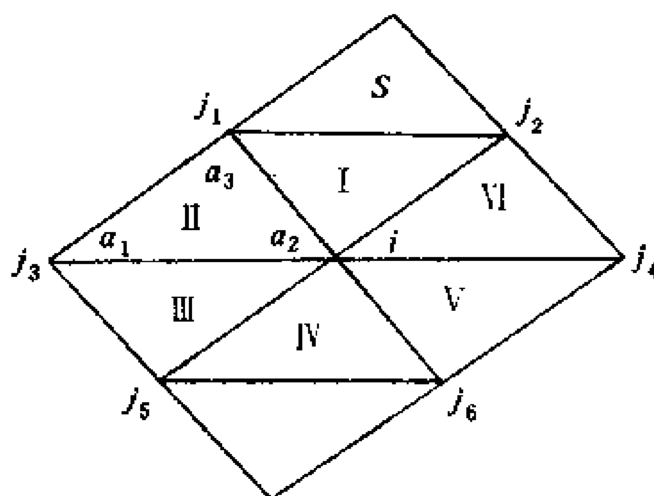


图 1.7

按如图 1.8 的分划那样采用坐标式编号，那么三角线性元离散格式就可写成下面形式：

$$[S_k]_{\triangle 1 k} = \begin{pmatrix} s_{-1,1} & s_{0,1} & 0 \\ s_{-1,0} & s_{0,0} & s_{1,0} \\ 0 & s_{0,-1} & s_{1,-1} \end{pmatrix}. \tag{1.20}$$

由第二篇第二章中关于三角线性元基函数公式 (2.12) 和总刚阵元素计算公式(2.27)～(2.31)，可知与编号为 i（是整体累加

编号)的节点有关的有限元方程,即有限元方程组中第 i 个方程为

$$\sum_{m=1}^{6} a(\varphi_i,\ \varphi_{j_m})u_{j_m}=(f,\ \varphi_i),$$

显然，把局部离散格式中心位置对准 i 节点，立刻就有对应关系

$$s_{-1,1}=a(\varphi_i,\ \varphi_{j_1})=-c_1,\quad s_{0,1}=a(\varphi_i,\ \varphi_{j_2})=-c_2,$$

$$s_{-1,0}=a(\varphi_i,\ \varphi_{j_3})=-c_3,\quad s_{0,0}=a(\varphi_i,\ \varphi_i)=2c,$$

$$(f,\ \varphi_i)=2Sf_i;$$

故一般**三角线性元局部离散格式**可形式地描述为

$$\begin{cases}[S_k]_{\Delta 1h}u_h(x)=F_{k\Delta 1h}(x), \quad \forall x\in \Omega_h,\\ [S_k]_{\Delta 1h}=\begin{pmatrix}-c_1 & -c_2 & 0\\ -c_3 & 2c & -c_3\\ 0 & -c_2 & -c_1\end{pmatrix},\\ F_{k\Delta 1h}(x)=2Sf(x).\end{cases} \tag{1.21}$$

正三角线性元格式是

$$\begin{cases}[S_k]_{\Delta 1h}u_h(x)=F_{k\Delta 1}(x),\ \forall x\in \Omega_h,\\ [S_k]_{\Delta 1}=\begin{pmatrix}-\sqrt{3}/3 & -\sqrt{3}/3 & 0\\ -\sqrt{3}/3 & 2\sqrt{3} & -\sqrt{3}/3\\ 0 & -\sqrt{3}/3 & -\sqrt{3}/3\end{pmatrix},\\ F_{k\Delta 1}(x)=2Sf(x) \text{ 或更高精度的数值积分.}\end{cases} \tag{1.22}$$

若采用直角三角形分划，并采用坐标式编号，参见图 1.8,

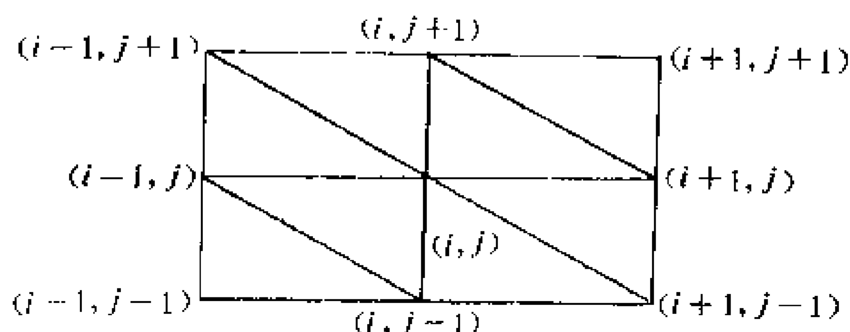


图 1.8

那么由（1.21）式立刻可得到在节点（i，j）处的局部**直角三**

角线性元格式：

$$
\begin{cases}
[S_k]_{\Delta 1 h} u_h(x) = F_{k\Delta 1 h}(x), \quad \forall x \in \Omega_h, \\
[S_k]_{\Delta 1 h} = \begin{pmatrix} 0 & -h_1^2 & 0 \\ -h_2^2 & 2(h_1^2+h_2^2) & -h_2^2 \\ 0 & -h_1^2 & 0 \end{pmatrix}, \\
F_{k\Delta 1 h}(x) = h_1^2 h_2^2 f(x).
\end{cases}
\tag{1.23}
$$

等腰直角三角线性元格式是：

$$
\begin{cases}
[S_k]_{\Delta 1} u_h(x) = F_{k\Delta 1}(x), \quad \forall x \in \Omega_h, \\
[S_k]_{\Delta 1} = \begin{pmatrix} 0 & -1 & 0 \\ -1 & 4 & -1 \\ 0 & -1 & 0 \end{pmatrix}, \\
F_{k\Delta 1}(x) = h^2 f(x).
\end{cases}
\tag{1.24}
$$

对于直角三角二次元，也可写出其相应的局部离散格式。如图 1.9 所示，设一个直角三角形单元 I 上的节点基函数为 $\varphi_k(k=1, 2, \cdots, 6)$. 其面积坐标表示形式参见第二篇第 4 章§3；

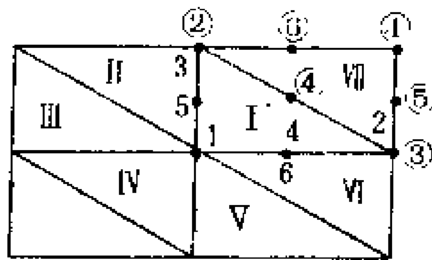


图 1.9

还记$\int_{\mathrm{I}} \cdot = a_{\mathrm{I}}(\cdot, \cdot)$ 表示总刚阵元素 $a(\cdot, \cdot)$ 在单元 I 上的相应积分，单元面积 $S = \frac{1}{2}h_1 h_2$，则可通过单元 I 上的有关积分计算，得到总刚阵中的相应元素．例如，在单元 I 上，利用第二篇中 (2.15) 式，有

$$
\int_{\mathrm{I}} \left[\left(\frac{\partial \varphi_4}{\partial x}\right)^2 + \left(\frac{\partial \varphi_4}{\partial y}\right)^2\right] dx dy
$$

$$
= \int_{\mathrm{I}} \left(4 L_3 \frac{\partial L_2}{\partial x} + 4 L_2 \frac{\partial L_3}{\partial x}\right)^2 dx\, dy
$$

$$+\int_{Ⅰ}\left(4L_3\frac{\partial L_2}{\partial y}+4L_2\frac{\partial L_3}{\partial y}\right)^2 dxdy$$

$$=16\frac{h_2}{h_1}\int_0^1\int_0^1 L_3^2 d\xi d\eta+16\frac{h_1}{h_2}\int_0^1\int_0^1 L_2^2 d\xi d\eta,$$

再利用 Euler 积分公式，就有

$$a_1(\varphi_4,\ \varphi_4)=\iint_{Ⅰ}\left[\left(\frac{\partial\varphi_4}{\partial x}\right)^2+\left(\frac{\partial\varphi_4}{\partial y}\right)^2\right]dxdy$$

$$=\frac{4}{3}\left(\frac{h_2}{h_1}+\frac{h_1}{h_2}\right);$$

用相同的办法计算得

$$a_1(\varphi_4,\ \varphi_i)=0,\ i=1,\ 2,\ 3,$$

$$a_1(\varphi_4,\ \varphi_5)=-\frac{4}{3}\ \frac{h_2}{h_1},\ a_1(\varphi_4,\ \varphi_6)=-\frac{4}{3}\ \frac{h_1}{h_2}.$$

再将单元Ⅶ的相应积分计算出来；且将节点“4”的支集上的所有积分综合，就得到与节点“4”有关的总刚阵元素和右端项元素：

$$\begin{cases}a(\varphi_4,\ \varphi_1)=a(\varphi_4,\ \varphi_{①})=0,\\ a(\varphi_4,\ \varphi_2)=a(\varphi_4,\ \varphi_3)=0,\\ a(\varphi_4,\ \varphi_5)=a(\varphi_4,\ \varphi_{⑤})=-\dfrac{4}{3}\ \dfrac{h_2}{h_1},\\ a(\varphi_4,\ \varphi_6)=a(\varphi_4,\ \varphi_{⑥})=-\dfrac{4}{3}\ \dfrac{h_1}{h_2},\\ a(\varphi_4,\ \varphi_4)=\dfrac{8}{3}\left(\dfrac{h_2}{h_1}+\dfrac{h_1}{h_2}\right),\\ (f,\ \varphi_4)=\dfrac{1}{3}h_1h_2f_4\ \text{或更高精度的数值积分。}\end{cases}$$

完全类似于节点“4”情形的办法，可相应计算出与节点“1”、“6”和“5”有关的总刚阵元素及右端向量元素。

最后，将局部离散格式的中心点分别与节点“1”、“6”、“5”和“4”对应，并利用前面求得的总刚阵元素和对称性，就得到下面的**直角三角二次元局部离散格式**：

$$\begin{cases} [S_k]^a_{\Delta 2h} u_h(x) = F^a_{k\Delta 2h}(x), \ x \text{ 是三角单元顶点}, \\ [S_k]^a_{\Delta 2h} = \dfrac{1}{3}\begin{bmatrix} & & h_1^2 & & \\ & & -4h_1^2 & & \\ h_2^2 & -4h_2^2 & 6(h_1^2+h_2^2) & -4h_2^2 & h_2^2 \\ & & -4h_1^2 & & \\ & & h_1^2 & & \end{bmatrix}, \\ F^a_{k\Delta 2h}(x) = 0 \ \text{ 或更高精度的数值积分}; \end{cases}$$

(1.25a)

$$\begin{cases} [S_k]^b_{\Delta 2h} u_h(x) = F^b_{k\Delta 2h}(x), \ x \text{ 是单元边中点}, \\ [S_k]^b_{\Delta 2h} = \dfrac{4}{3}\begin{bmatrix} & -h_1^2 & \\ -h_2^2 & 2(h_1^2+h_2^2) & -h_2^2 \\ & -h_1^2 & \end{bmatrix}, \\ F^b_{k\Delta 2h}(x) = \dfrac{1}{3} h_1^2 h_2^2 f(x) \text{或更高精度的数值积分}. \end{cases} \tag{1.25b}$$

令 $h_1 = h_2 = h$，即得到**等腰直角三角二次元格式**：

$$\begin{cases} [S_k]^a_{\Delta 2} u_h(x) = F^a_{k\Delta 2}(x), \ x \text{是三角单元顶节点}, \\ [S_k]^a_{\Delta 2} = \dfrac{1}{3}\begin{bmatrix} & & 1 & & \\ & & -4 & & \\ 1 & -4 & 12 & -4 & 1 \\ & & -4 & & \\ & & 1 & & \end{bmatrix}, \\ F^a_{k\Delta 2}(x) = 0 \text{ 或更高精度的数值积分}; \end{cases} \tag{1.26a}$$

$$\begin{cases} [S_k]^b_{\Delta 2} u_h(x) = F^b_{k\Delta 2}(x), \ x \text{ 是单元边中点}, \\ [S_k]^b_{\Delta 2} = \dfrac{4}{3}\begin{bmatrix} & -1 & \\ -1 & 4 & -1 \\ & -1 & \end{bmatrix}, \\ F^b_{k\Delta 2}(x) = \dfrac{1}{3} h^2 f(x) \ \text{ 或更高精度的数值积分}. \end{cases} \tag{1.26b}$$

由第二篇第四章的（4.12）式和（4.14）式也容易写出一般**矩形双线性元局部离散格式**：

$$
\left\{
\begin{aligned}
&[S_k]_{\square 1h} u_h(x) = F_{k\square 1h}(x), \quad \forall x \in \Omega_h, \\
&[S_k]_{\square 1h} = \begin{pmatrix}
-\frac{1}{6}(h_1^2 + h_2^2) & \frac{1}{3}h_2^2 - \frac{2}{3}h_1^2 & -\frac{1}{6}(h_1^2 + h_2^2) \\
\frac{1}{3}h_1^2 - \frac{2}{3}h_2^2 & \frac{4}{3}(h_1^2 + h_2^2) & \frac{1}{3}h_1^2 - \frac{2}{3}h_2^2 \\
-\frac{1}{6}(h_1^2 + h_2^2) & \frac{1}{3}h_2^2 - \frac{2}{3}h_1^2 & -\frac{1}{6}(h_1^2 + h_2^2)
\end{pmatrix}, \\
&F_{k\square 1h}(x) = h_1^2 h_2^2\left(1 - \frac{1}{6}[S_k]_5\right)f(x).
\end{aligned}
\right.
\tag{1.27}
$$

正方形双线性元格式是：

$$
\left\{
\begin{aligned}
&[S_k]_{\square 1} u(x) = F_{k\square 1}(x), \qquad \forall x \in \Omega_h, \\
&[S_k]_{\square 1} = \begin{pmatrix}
-1/3 & -1/3 & -1/3 \\
-1/3 & 8/3 & -1/3 \\
-1/3 & -1/3 & -1/3
\end{pmatrix}, \\
&F_{k\square 1} = h^2\left(1 - \frac{1}{6}[S_k]_5\right)f(x).
\end{aligned}
\right.
\tag{1.28}
$$

矩形双二次元局部离散格式是

$$
\left\{
\begin{aligned}
&[S_k]^a_{\square 2h} u_h(x) = F^a_{k\square 2h}(x), \quad x\text{ 是单元顶点}, \\
&[S_k]^a_{\square 2h} = \frac{1}{45}\begin{pmatrix}
H & Q & E & Q & H \\
Q & N & D & N & Q \\
A & B & C & B & A \\
Q & N & D & N & Q \\
H & Q & E & Q & E
\end{pmatrix}, \\
&F^a_{k\square 2h}(x) = \frac{1}{9}\left(1 - \frac{1}{6}[S_k]_5\right)f(x),
\end{aligned}
\right.
\tag{1.29a}
$$

其中，$A = 4h_2^2 - 7h_1^2$，$B = -32h_2^2 + 14h_1^2$，$C = 56(h_1^2 + h_2^2)$，
$D = 14h_2^2 - 32h_1^2$，$E = -7h_2^2 + 4h_1^2$，$N = -8(h_1^2 + h_2^2)$，

$Q=4h_2^2+h_1^2$, $H=-\frac{1}{2}(h_1^2+h_2^2)$;

$$\begin{cases}[S_k]_{\square 2h}^{b}u_h(x)=F_{k\square 2h}^{b}(x),\ x\ \text{是单元横向边中点},\\[S_k]_{\square 2h}^{b}=\dfrac{1}{45}\begin{vmatrix}4h_2^2+h_1^2 & -8h_2^2+8h_1^2 & 4h_2^2+h_1^2\\-8(h_2^2+h_1^2) & 16h_2^2-64h_1^2 & -8(h_2^2+h_1^2)\\-32h_2^2+14h_1^2 & 64h_2^2+112h_1^2 & -32h_2^2+14h_1^2\\-8(h_2^2+h_1^2) & 16h_2^2-64h_1^2 & -8(h_2^2+h_1^2)\\4h_2^2+h_1^2 & -8h_2^2+8h_1^2 & 4h_2^2+h_1^2\end{vmatrix},\\F_{k\square 2h}^{b}(x)=\dfrac{2}{9}h_1^2h_2^2\left(1-\dfrac{1}{6}[S_k]_5\right)f(x);\end{cases}\tag{1.29b}$$

$$\begin{cases}[S_k]_{\square 2h}^{c}u_h(x)=F_{k\square 2h}^{c}(x),\ x\ \text{是单元纵向边中点},\\[S_k]_{\square 2h}^{c}\ \text{是}\ [S_k]_{\square 2k}^{b}\ \text{的转置形式},\\F_{k\square 2h}^{c}(x)=F_{k\square 2h}^{b}(x);\end{cases}\tag{1.29c}$$

$$\begin{cases}[S_k]_{\square 2h}^{d}u_h(x)=F_{k\square 2h}^{d}(x),\ x\ \text{是单元中心点},\\[S_k]_{\square 2h}^{d}=\dfrac{8}{45}\begin{vmatrix}-h_2^2+h_1^2 & 2h_2^2-8h_1^2 & -h_2^2+h_1^2\\-8h_2^2+2h_1^2 & 18(h_2^2+h_1^2) & -8h_2^2+2h_1^2\\-h_2^2+h_1^2 & 2h_2^2-8h_1^2 & -h_2^2+h_1^2\end{vmatrix},\\F_{k\square 2h}^{d}(x)=\dfrac{4}{9}h_1^2h_2^2\left(1-\dfrac{1}{6}[S_k]_5\right)f(x).\end{cases}\tag{1.29d}$$

对于**正方形双二次元格式**，只要令 $h_1=h_2=h$，即可从(1.29a)～(1.29d) 诸式中得到：

$$\begin{cases}[S_k]_{\square 2}^{a}u_h(x)=F_{k\square 2}^{a}(x),\ x\ \text{是单元顶点},\\[S_k]_{\square 2}^{a}=\dfrac{1}{45}\begin{pmatrix}-1 & 5 & -3 & 5 & -1\\5 & -16 & -18 & -16 & 5\\-3 & -18 & 112 & -18 & -3\\5 & -16 & -18 & -16 & 5\\-1 & 5 & -3 & 5 & -1\end{pmatrix},\\F_{k\square 2}^{a}(x)=\dfrac{1}{9}h^2\left(1-\dfrac{1}{6}[S_k]_5\right)f(x);\end{cases}\tag{1.30a}$$

$$
\begin{cases}
[S_k]^b_{\square 2} u_h(x) = F^b_{k\square 2}(x), \quad x\text{是单元横向边中点}, \\
[S_k]^b_{\square 2} = -\dfrac{1}{45}\begin{pmatrix} 5 & 0 & 5 \\ -16 & -48 & -16 \\ -18 & 176 & -18 \\ -16 & -48 & -16 \\ 5 & 0 & 5 \end{pmatrix}, \\
F^b_{k\square 2}(x) = \dfrac{2}{9}h^2\left(1 - \dfrac{1}{6}[S_k]_6\right)f(x);
\end{cases}
\tag{1.30b}
$$

$$
\begin{cases}
[S_k]^c_{\square 2} u_h(x) = F^c_{k\square 2}(x), \quad x\text{是单元纵向边中点}, \\
[S_k]^c_{\square 2} = \dfrac{1}{45}\begin{pmatrix} 5 & -16 & -18 & -16 & 5 \\ 0 & -48 & 176 & -48 & 0 \\ 5 & -16 & -18 & -16 & 5 \end{pmatrix}, \\
F^c_{k\square 2}(x) = \dfrac{2}{9}h^2\left(1 - \dfrac{1}{6}[S_k]_6\right)f(x);
\end{cases}
\tag{1.30c}
$$

$$
\begin{cases}
[S_k]^d_{\square 2} u_h(x) = F^d_{k\square 2}(x), \quad x\text{ 是单元中心点}, \\
[S_k]^d_{\square 2} u_h(x) = \dfrac{16}{45}\begin{pmatrix} 0 & -3 & 0 \\ -3 & 16 & -3 \\ 0 & -3 & 0 \end{pmatrix}, \\
F^d_{k\square 2}(x) = \dfrac{4}{9}h^2\left(1 - \dfrac{1}{6}[S_k]_6\right)f(x).
\end{cases}
\tag{1.30d}
$$

最后，我们指出，上述有限元局部离散格式也可以由几个单元刚度矩阵叠加而得到，读者可自行验证之。

2.2 差分格式和有限元格式的某些统一性

若比较（1.14）式和（1.23）式、（1.15）式和（1.24)式，可知5点差分格式完全同于直角三角线性元格式，即

$$
[S_k]_{5h} = [S_k]_{\Delta 1h}, \qquad [S_k]_h = [S_k]_{\Delta 1}. \tag{1.31}
$$

若比较（1.16）式和（1.27）式、（1.17）式和（1.28)式，

可知 9 点差分格式是直角三角线性元格式和矩形双线性元格式的平均，即

$$
\begin{cases}
[S_k]_{9h} = \dfrac{1}{2}([S_k]_{\Delta 1h} + [S_k]_{\square 1h}), \\
[S_k]_9 = \dfrac{1}{2}([S_k]_{\Delta 1} + [S_k]_{\square 1}).
\end{cases} \tag{1.32}
$$

引用记号 $[S_k]_{\Delta 1h}$ 和 $[S_k]_{\Delta 1\frac{h}{2}}$ 分别表示在步长为 h 的网格和步长为 $h/2$ 的网格上定义在某节点处的直角三角线性元格式，记号 $[S_k]_{5h}$ 和 $[S_k]_{5\frac{h}{2}}$ 分别表示作用于某节点处的 h 和 $h/2$ 的网格上的 5 点差分格式. 比较（1.18）式、（1.14）式、（1.25a）和（1.25b）式就可看出，特殊 9 点差分格式也可看作三角二次元格式和三角线性元格式的线性组合，即

$$
[S_k]_{9h*} = 2(2[S_k]_{\Delta 2\frac{h}{2}} - 3[S_k]^{a}_{\Delta 2h}). \tag{1.33}
$$

直角三角二次元格式也是几种简单离散格式的线性组合，即

$$
\begin{cases}
[S_k]^{a}_{\Delta 2h} = \dfrac{1}{3}\left(2[S_k]_{\Delta 1\frac{h}{2}} - \dfrac{1}{2}[S_k]_{9h*}\right), \\
[S_k]^{a}_{\Delta 2h} = \dfrac{1}{3}(4[S_k]_{\Delta 1\frac{h}{2}} - [S_k]_{\Delta 1h}), \\
[S_k]^{b}_{\Delta 2h} = \dfrac{4}{3}[S_k]_{\Delta 1\frac{h}{2}}.
\end{cases} \tag{1.34}
$$

同样，矩形双二次元也一定能分解为一些简单格式的线性组合，这一工作留给读者自己去完成.

(1.31)～(1.34)式说明了解 Poisson 方程的差分方法和有限元方法在计算格式方面的统一性. 同样地，对一阶导数的模拟，这两类方法的计算格式也存在线性组合关系. 这种计算格式的统一性还说明了如下几个问题.

第一，局部离散格式用于实际计算是方便的. 有限元格式和

差分格式都是一种局部离散格式，它们都能方便地用于迭代法中。即使定解区域Ω不规则，一般也可以使用矩形网格剖分；对于正则内点使用规则的局部离散格式，此时无需复杂计算和存贮有关的总刚阵元素，迭代求解过程和程序设计都很简单；对于非正则内点可使用不规则的局部离散格式，此时要存贮少量必要的信息，迭代求解过程也很简单；因此，使用局部离散格式不仅可将有限元方法和差分方法在一般实际计算中统一起来，而且对减少计算量和存贮量都是大有好处的。

第二，多种低精度局部离散格式的组合应用可求得较高精度的局部离散格式。

例如，直角三角线性元格式、矩形双线性元格式、特殊 9 点差分格式和 5 点差分格式都是 2 阶精度的局部离散格式，它们的适当线性组合（1.32）式和（1.34）式就是 4 阶精度的局部离散格式。

又例如，若设求解 Poisson 方程的两个离散方程为

$$[S_k]_a u(x,y) = \sum_{i=1}^{m_1} V_i(f) h^{\alpha_i} + h^{\beta} B(u) + O(h^{\gamma}),$$

$$[S_k]_b u(x,y) = \sum_{j=1}^{m_2} W_j(f) h^{\alpha_j} + c h^{\beta} B(u) + O(h^{\gamma}),$$

其中 $[S_k]_a$ 和 $[S_k]_b$ 是由常数因子构成的 2 阶精度的局部离散格式，$V_i(f)$和$W_j(f)$为依赖于$f(x, y)$的算子，$B(u)$ 为依赖于 u 而不依赖于 h 的算子，c、α_l、β、γ、m_1 和 m_2 为实数，且 $\gamma>\beta>\alpha$，[illegible]$\max\{\alpha_{m_1}, \alpha_{m_2}\}$，则它们的线性组合

$$c[S_k]_a - [S_k]_b = c\sum_{i=1}^{m_1} V_i(f) h^{\alpha_i} - \sum_{j=1}^{m_2} W_j(f) h^{\alpha_j}$$

是逼近 Poisson 方程的且截断误差为$O(h^{\gamma-2})$阶精度的离散方程（参见习题29）。

复杂的高精度的离散微分方程是由简单的低精度离散方程组合而成的，这一事实将有利于校正算法、分裂算法和并行算法的

设计。

第三，椭圆问题的差分解也是弱解。在弱条件下获得的有限元局部离散格式和较高光滑性条件下获得的差分局部离散格式的统一性说明，在椭圆型方程的解光滑性不够高的情况下，如方程右端项、边值函数光滑性不够高、凸定解区域有些角点等情形，差分格式仍是行之有效的，其实际计算效果与有限元方法的效果是在某种程度上一致的。这一事实还表明，差分解和有限元解在解的结构方面应存在着某种固有联系，目前研究已发现有限元解和差分解都存在着共同形态的渐近展开式[22~25, 37]，这从一个方面反映了这一内在联系。其它的内在联系，还有待进一步研究。

第二章　基本迭代解法及其收敛性分析

在用局部离散格式表述差分方程和有限元方程的基础上，本章不仅介绍了解椭圆离散方程的诸种常用的基本迭代方法，而且用局部 Fourier 分析方法讨论了这些迭代方法的收敛速度，还对这些基本迭代解法作了比较，并讨论了迭代控制和迭代组合等应用问题。

§1　基本概念

不论采用差分方法还是采用有限元方法，线性椭圆型方程的离散形式都可统一写为形如

$$L_h u_h = f_h \tag{2.1}$$

的线性代数方程组，其中 $L_h=(a_{ij})_{N\times N}$ 非奇异，$L_h=L_h^l+L_h^d+L_h^r$，L_h^l、L_h^d 和 L_h^r 分别是 L_h 的下三角、对角和上三角矩阵，$u_h=(u_1,\ u_2,\ \cdots,\ u_N)^T$ 和 $f_h=(f_1,\ f_2,\ \cdots,\ f_N)^T$ 是离散方程的数值解和右端项函数在所有节点处离散值所构成的向量，N 是离散区域的总节点数。

u_h 的 Euclid 模、最大模和 l 模分别记为

$$\|u_h\|=\left(\sum_{i=1}^N u_i^2\right)^{1/2},\qquad \|u_h\|_\infty=\max_{1\le i\le N}|u_i|,$$

$$\|u_h\|_l=\sum_{i=1}^N |u_i|.$$

相应地，矩阵 L_h 的 Frobenius 模、行模和列模记为

$$\|L_h\|_F=\left(\sum_{i,j=1}^N a_{ij}^2\right)^{1/2},\qquad \|L_h\|_l=\max_{1\le i\le N}\sum_{j=1}^N|a_{ij}|,$$

$$\|L_h\|_r=\max_{1\le j\le N}\sum_{i=1}^N|a_{ij}|;$$

L_h 的 Euclid 模记为

$$\|L_h\| = (\rho(L_hL_h^*))^{1/2},$$

其中 L_h^*是 L_h 的共轭矩阵，$\rho(M)$是M阵的谱半径。矩阵 L_h 的谱半径不超过任一种其它的模，并且当 L_h 是对称阵时，有

$$\|L_h\| = \rho(L_h).$$

若能将线性代数方程组(2.1)改写成

$$u_h = T_hu_h + g_h \tag{2.2}$$

的形式，且其迭代公式为

$$u_h^{k+1} = T_hu_h^k + g_h, \tag{2.3}$$

式中 T_h为迭代矩阵，u_h^k表示初始值 u_h^0 经 k 次迭代后的值。形如(2.3)式的迭代法，我们统称为基本迭代法。若基本迭代法收敛，则(2.3)式近似解的极限必是(2.2)式的解。

定理2.1 解方程组(2.2)的迭代过程(2.3)对任意右端项 g_h 及任意初始向量 u_h^0收敛的充要条件是 $\rho(T_h)<1$。

下面讨论基本迭代法的收敛速度问题。设迭代矩阵 T_h 对称且有N个特征值 λ_1，λ_2，…，λ_N，相应的相互正交的特征函数为 φ_1，φ_2，…，φ_N，即有

$$T_h\varphi_j = \lambda_j\varphi_j, \quad j = 1, 2, \cdots, N;$$

则对任意初始值u_h^0，有

$$u_h - u_h^0 = \sum_{j=1}^{N} a_j\varphi_j.$$

由此，并利用

$$u_h - u_h^k = T_h(u_h - u_h^{k-1}) = T_h^k(u_h - u_h^0),$$

就得到

$$u_h - u_h^k = \sum_{j=1}^{N} \lambda_j^k a_j\varphi_j,$$

从而由 φ_j 的正交性得到

$$\|u_h - u_h^k\| \leqslant \rho^k(T_h)\|u_h - u_h^0\|.$$

由上式看出，$\rho(T_h)$的大小决定了迭代收敛的快慢，例如要使初始误差经 k 步迭代后被压缩 $1/e$ 倍，即要求

$$\rho^k(T_h)\|u_h - u_h^0\| \leqslant e^{-1}\|u_h - u_h^0\|,$$

将上式两边取对数，就得到所需迭代步数的估计 $k \geq -1/\ln\rho(T_h)$。称 $R(T_h) = -\ln\rho(T_h)$ 为迭代法的**收敛速度**。收敛速度的意义是明显的，即 $1/R(T_h)$ 表示了使初始误差压缩为原来误差 $1/e$ 倍所需的最小迭代步数。

若 T_h 是任意矩阵，相应地应有

$$\|u_h - u_h^k\| \leqslant \|T_h^k\| \|u_h - u_h^0\|,$$

此时的收敛速度应定义为

$$R_k(T_h) = -\ln\|T_h^k\|^{1/k} = -\frac{1}{k}\ln\|T_h^k\|,$$

$R_k(T_h)$称为迭代法的**平均收敛速度**。

由于对任意矩阵 T_h，均有

$$\lim_{k\to\infty} R_k(T_h) = -\ln\rho(T_h),$$

因此，总可以用 $R(T_h) = -\ln\rho(T_h)$表示迭代法的收敛速度，它又称为**渐近收敛速度**。

§2 局部Fourier分析法

本节介绍一种**局部 Fourier 分析方法**，用这种方法可以很方便地分析离散格式的特征值问题，也可以很方便地分析某种迭代方法的收敛速度问题，有兴趣的读者还可参阅文献[12]。

先讨论 L_h 的特征值问题。

设对定解区域[0，1]×[0，1]作矩形分划，网将参数为 h_1 和 h_2，那么在一切网格节点(i, j)处给定的离散值 u_{ij} 可扩展定义为一个确定在 Ω_h 上的函数 $u_h(x_1, x_2)$，例如它可以是阶梯函数。若微分方程的解可以周期延拓到整个实(复)平面上（例如矩形定解区域及第一边值条件或其它局期边界条件的情形等），则 $u_h(x_1, x_2)$ 也可以周期地延拓到**无限网**将 G_h 上，且可展成 Fourier 级数

$$u_h(x_1, x_2) = \sum_{l_1, l_2=-\infty}^{+\infty} v(l_1, l_2)\exp[i\pi(l_1, l_2)(x_1, x_2)], \tag{2.4}$$

式中，$v(l_1, l_2)$是 Fourier 系数，且规定$(l_1, l_2)(x_1, x_2)=(l_1x_1, l_2x_2)$。

设 9 点离散格式

$$L_h=[S_k]=\begin{pmatrix} s_{-1,1} & s_{0,1} & s_{1,1} \\ s_{-1,0} & s_{0,0} & s_{1,0} \\ s_{-1,-1} & s_{0,-1} & s_{1,-1} \end{pmatrix} \tag{2.5}$$

也具有延拓性，那么由定义在 G_h 上的离散微分方程可统一写为

$$\begin{aligned} L_hu_h(x_1, x_2) &= [S_k]u_h(x_1, x_2) \\ &= \sum_{k\in V} s_k u_h((x_1, x_2)+(k_1, k_2)(h_1, h_2)) \\ &= \tilde{f}(x_1, x_2), \quad \forall (x_1, x_2)\in G_h \end{aligned}$$

的形式，式中$[S_k]$和 $\tilde{f}$ 应随不同离散化方式而异(参见第一章)，$k=(k_1, k_2)$，V 表示离散格式中所涉及的(k_1, k_2)的集合。因此，可利用 Fourier 分析方法研究具有周期延拓性的 9 点离散格式的特征值。

为此，设 $\lambda(h, k)$和 $\varphi(h, x)$($x=(x_1, x_2)\in G_h$) 是 L_h 的特征值和相应的特征函数，将形如(2.4)式的 $\varphi(h, x)$的Fourier 展式代入

$$L_h\varphi(h, x)=\lambda(h, k)\varphi(h, x), \quad \forall x\in G_h,$$

就有

$$\begin{aligned} &\sum_{k\in V} s_k \sum_{l_1, l_2=-\infty}^{+\infty} v(l_1, l_2)\exp[i\pi(l_1, l_2)((x_1, x_2) \\ &\qquad +(k_1, k_2)(h_1, h_2))] \\ &= \lambda(h, k)\sum_{l_1, l_2=-\infty}^{+\infty} v(l_1, l_2)\exp[i\pi(l_1, l_2)(x_1, x_2)], \end{aligned}$$

把上式左端写成

$$\sum_{l_1, l_2=-\infty}^{+\infty} v(l_1, l_2)\exp[i\pi(l_1, l_2)(x_1, x_2)] \cdot \sum_{k\in V} s_k\exp[i\pi(l_1, l_2)(h_1k_1, h_2k_2)]$$

的形式，并在等式两边约去公因子，就有

$$\lambda(h,\ \mathrm{k})=\sum_{\mathrm{k}\in V}s_{\mathrm{k}}\exp[i\pi(l_1,\ l_2)(h_1\mathrm{k}_1,\ h_2\mathrm{k}_2)].$$

若令 $\theta=(\theta_1,\ \theta_2)=(l_1h_1,\ l_2h_2)\pi$, $|\theta|=\max\{|\theta_1|,\ |\theta_2|\}$,则$L_h$的特征值可表示为

$$\lambda(\theta,\ \mathrm{k})=\sum_{\mathrm{k}\in V}s_{\mathrm{k}}\exp(i\theta\mathrm{k}),\qquad |\theta|\leqslant\pi. \tag{2.6}$$

请读者注意 θ 的定义域. 如果 $u_h(x_1,\ x_2)$定义在无限网格上, 此时 $l\in(-\infty,\ +\infty)$, 因为 $\exp[i(n\pi+\theta)]=\pm\exp(i\theta)$, 所以可以取 $-\pi\leqslant\theta\leqslant\pi$, 且当 $\theta\to 0$ 时, $|\lambda(\theta,\ \mathrm{k})|$取得极大值. 如果将 $u_h(x_1,\ x_2)$的 Fourier 展式限制在有限的矩形网格 Ω_h 上, 那么(2.6)式就可表示在有限维网格函数空间中求局部离散格式L_h的特征值; 此时 $l_1\in\{1,\ 2,\ \cdots,\ N_1\}$, $l_2\in\{1,\ 2,\cdots,\ N_2\}$, $|\theta|\in(0,\ \pi)$, 且$|\theta|=\max\{|\theta_1|,\ |\theta_2|\}=O(h)$; 当 $\theta=(\pi h_1,\ \pi h_2)$时, $|\lambda(\theta,\ \mathrm{k})|$取得最大值.

利用(2.6)式, 很容易求出 5 点离散格式 $[S_{\mathrm{k}}]_5$的特征值

$$\lambda(\theta,\ \mathrm{k})=4-\exp(i\theta_1)-\exp(-i\theta_1)-\exp(i\theta_2)-\exp(-i\theta_2)$$

$$=4-2\cos\theta_1-2\cos\theta_2=4\sin^2\frac{\theta_1}{2}+4\sin^2\frac{\theta_2}{2},$$

和谱半径

$$\rho([S_{\mathrm{k}}]_5)=O(h^2).$$

同样易知 9 点离散格式的谱半径也是 $O(h^2)$.

再讨论松驰迭代矩阵的特征值表示问题.

首先值得注意的是, 由于差分格式和有限元格式的局部性, 松驰迭代过程也是局部发生改变的. 也就是说, 每个完整的松驰迭代步实际上是由若干个局部迭代过程组合的结果. 事实上, 要改变$(i,\ j)$节点处的近似值, 松驰迭代过程只涉及到该节点处和与该节点相邻的那些节点处的近似值, 松驰迭代过程中所用到的节点指标集与局部离散格式的指标集V相同, 所以又可以称V为**格式点集**. 显然, 在松驰迭代过程中格式点集V可分解为

$$V=V^+\cup V^0\cup V^-,$$

其中，V^+表示那些在迭代过程中要用新值代替旧值的网格点集合；

V^0表示在迭代过程中需要求出新值的那些网格点集合；

V^-表示在迭代过程中使用原(旧)值的那些网格点集合。

与格式点集分解相应，离散格式也可分解为

$$[S_k] = [S_k^+] + [S_k^0] + [S_k^-];$$

$[S_k]$的特征值 $\lambda(\theta, h)$(这种写法与 $\lambda(\theta, k)$ 一样，是为了醒目地突出参变量）也可相应地分解为

$$\lambda(\theta, h) = \lambda^+(\theta, h) + \lambda^0(\theta, h) + \lambda^-(\theta, h),$$

式中 λ^+、λ^0 和 λ^-分别是 L_h^+、L_h^0和L_h^-的特征值。

若设 v_h 和 $\bar{v}_h$ 分别表示松弛迭代步前后的误差，则松弛迭代过程（参见本章中提到的各种松弛方法，特别参见 SOR 方法的矩阵表示式）一般可描述为

$$\begin{cases}([S_k^0] + \omega[S_k^+])\bar{v}_h(x) = ((1-\omega)[S_k^0] - \omega[S_k^-])v_h(x), & x \in \tilde{G}_h, \\ \bar{v}_h(x) = v_h(x), & x \in G_h \backslash \tilde{G}_h, \end{cases} \tag{2.7}$$

式中，$\tilde{G}_h$ 是 G_h 的子集，它表示松弛迭代过程只是在 $\tilde{G}_h$ 中的那些节点处施行。

设 $\varphi(\theta, x)$是松弛迭代矩阵 T_h 的特征向量，$A(\theta, h, \omega)$ 是其对应的特征值，即有

$$T_h\varphi(\theta, x) = A(\theta, h, \omega)\varphi(\theta, x), \tag{2.8}$$

那么利用(2.6)～(2.8)式就很容易确定迭代矩阵特征值 $A(\theta, h, \omega)$的表达式。总上所述，我们有

定理2.2 形如(2.5)式的离散格式的特征值由(2.6)式确定；由(2.7)式定义的松弛迭代过程的特征值由下式确定：

$$A(\theta, h, \omega) = \frac{(1-\omega)\lambda^0(\theta, h) - \omega\lambda^-(\theta, h)}{\lambda^0(\theta, h) + \omega\lambda^+(\theta, h)}. \tag{2.9}$$

值得指出的是，定理 2.2 中的结论是在微分方程解和离散格式都能在整个实(复)平面上作周期延拓的条件下获得的，此时定

理2.2中的结论是准确的。如果微分方程的解不能从有限区域周期延拓出去，则可认为离散格式定义在无限网格上，此时定理中所述结论只能近似描述迭代矩阵的特征值，但仍能较好地描述远离边界的那些内点处局部迭代过程的特性。实践表明：对于凸区域上的椭圆型第一边值问题，由于强制边界条件的影响大，靠近边界点处的迭代数值解不仅精度较高而且收敛较快；由于强制边界条件的影响小，远离边界点处的迭代数值解精度偏低且收敛较慢。因此，局部Fourier分析方法可以较好地反应迭代过程的整体收敛特性。

§3 ω-Jacobi迭代法

若网格节点采用整体累加规则编号，编号从1到N，固定某个i，则解离散方程(2.1)的ω-Jacobi**迭代公式**为

$$u_i^{k+1}=u_i^k+\frac{\omega}{a_{ii}}\Big(f_i-\sum_{j=1}^{N}a_{ij}u_j^k\Big).$$

现在迭代过程中进一步考虑离散格式和松弛过程的局部特征，此时ω-Jacobi迭代的格式点集分解情形为

$$V^+=\varnothing,\quad V^0=\{i\},\quad V^-=\{j\mid j\text{ 是 }i\text{ 的邻点}\},$$

于是ω-Jacobi迭代公式又可改写为

$$u_i^{k+1}=(1-\omega)u_i^k+\frac{\omega}{a_{ii}}\Big(f_i-\sum_{j\in V^-}a_{ij}u_j^k\Big),\qquad(2.10)$$

此式可用于不规则网格分划（如一般三角形剖分等）的离散方程求解。

对直角三角形剖分和矩形网格剖分，网格节点可按坐标形式编号。当采用9点局部离散格式时，ω-Jacobi迭代格式点集V及其分解情形为

$V=\{(k_1,k_2)\mid k_1,k_2=0,\ \pm1\}$，　$V^+=\varnothing$，$V^0=\{(0,0)\}$，

$V^-=\{(k_1,k_2)\mid k_1,k_2=0,\ \pm1,\ k_1$和$k_2$不同时为零$\}$；

或简记为符号格式

$$V=\begin{pmatrix} - & - & - \\ - & 0 & - \\ - & - & - \end{pmatrix}.$$

因此，利用离散格式和符号格式很容易写出节点按坐标编号情况下的 ω-Jacobi 迭代公式：

$$u_{ij}^{k+1}=(1-\omega)u_{ij}^{k}+\frac{\omega}{s_{0,0}}(f_{ij}-s_{-1,1}u_{i-1,j+1}^{k}-s_{0,1}u_{i,j+1}^{k}$$

$$-s_{1,1}u_{i+1,j+1}^{k}-s_{-1,0}u_{i-1,j}^{k}-s_{1,0}u_{i+1,j}^{k}-s_{-1,-1}u_{i-1,j-1}^{k}$$

$$-s_{0,-1}u_{i,j-1}^{k}-s_{1,-1}u_{i+1,j-1}^{k}). \tag{2.11}$$

用此式计算，每个完整的迭代步至多需要 $10N$ 次乘(除)法和 $10N$ 次加(减)法运算，实际计算时，一般需要两组存贮单元，以存放 u_h^k 和 u_h^{k+1}。

利用局部 Fourier 分析方法，可分析 ω-Jacobi 迭代的收敛性。例如对 5 点离散格式 $[S_k]_5$，运用定理 2.2 中结论有

$$[S_k]_5=\begin{pmatrix} & -1 & \\ -1 & 4 & -1 \\ & -1 & \end{pmatrix},\quad \begin{aligned}&\lambda^0=4,\quad \lambda^+=0,\\ &\lambda^-=-\exp(i\theta_1)-\exp(i\theta_2)\\ &\qquad -\exp(-i\theta_1)-\exp(-i\theta_2),\end{aligned}$$

故
$$\begin{aligned}A(\theta,h,\omega)&=\frac{1}{4}[4(1-\omega)+\omega(2\cos\theta_1+2\cos\theta_2)]\\ &=1-\frac{\omega}{2}(2-\cos\theta_1-\cos\theta_2)\\ &=1-\omega\left(\sin^2\frac{\theta_1}{2}+\sin^2\frac{\theta_2}{2}\right);\end{aligned}$$

只需令 $\theta_1=\theta_2=\pi h$，就有

$$\rho(T_h)=\begin{cases}\left|1-2\omega\sin^2\dfrac{\pi h}{2}\right|\approx 1-\dfrac{\omega\pi^2h^2}{2}\\ =1-O(h^2), \qquad 0<\omega\leqslant 1,\\ \geqslant 1, \quad \omega\leqslant 0\text{或}\omega>1.\end{cases}$$

对于 9 点差分格式 $[S_h]_9$ 情形，应用定理 2.2 中的结论有

$$
\begin{aligned}
&\lambda^0 = 20, \qquad \lambda^+ = 0, \\
&\lambda^- = -\exp[i(-\theta_1+\theta_2)] - \exp[i(-\theta_1-\theta_2)] \\
&\qquad - \exp[i(\theta_1+\theta_2)] - \exp[i(\theta_1-\theta_2)] \\
&\qquad - 4[\exp(-i\theta_1) + \exp(i\theta_1) + \exp(-i\theta_2) + \exp(i\theta_2)], \\
&A(\theta, h, \omega) = (1-\omega) + \frac{\omega}{5}(\cos\theta_1\cos\theta_2 + 2\cos\theta_1 + 2\cos\theta_2);
\end{aligned}
$$

用 $\theta_2 = \theta_1 = \pi h$，$\cos\theta_1 \approx 1 - \theta_1^2/2$ 代入上式，得

$$
\rho(T_h) = \begin{cases} 1 - \dfrac{3}{5}\omega\pi^2 h^2 = 1 - O(h^2), & 0 < \omega \leqslant 1, \\ \geqslant 1, & \omega \leqslant 0 \ \text{或} \ \omega > 1. \end{cases}
$$

对于正方形双线性格式 $[S_h]_{\square 1}$ 情形，同样可算出

$$
\rho(T_h) = \begin{cases} 1 - \dfrac{3\omega}{4}\pi^2 h^2 = 1 - O(h^2), & 0 < \omega \leqslant 1, \\ \geqslant 1, & \omega \leqslant 0 \text{或} \omega > 1. \end{cases}
$$

总之，上述离散格式的 ω-Jacobi 迭代的谱半径虽有差别，但其迭代收敛速度都与 h^2 同阶，换言之，用 ω-Jacobi 方法解上述离散方程总共需要的迭代步数与 $h^{-2} = N^2$ 同阶；另外，当 $\omega = 1$ 时，ω-Jacobi 迭代方法收敛最快。

为了便于分析，也常用到 ω-Jacobi 迭代的矩阵表示式

$$
\begin{aligned}
u_h^{k+1} &= u_h^k + \omega(L_h^d)^{-1}(f_h - L_h u_h^k) \\
&= (I - \omega(L_h^d)^{-1}L_h)u_h^k + \omega(L_h^d)^{-1}f_h, \qquad (2.12)
\end{aligned}
$$

其迭代矩阵为 $I - \omega(L_h^d)^{-1}L_h$。

§4 GS迭代法

Gauss-Seidel 迭代法(简称**GS迭代法**)修正了Jacobi 方法。事实上，从公式(2.10)和(2.12)可知，在 Jacobi 迭代过程的每

一步计算中，只是利用 u_h^k 的分量来计算 u_h^{k+1} 的分量，在计算 u_h^{k+1} 的第 i 个分量 $u_i^{k+1}(i>1)$ 时，对已经计算出的最新分量 u_1^{k+1}，…，u_{i-1}^{k+1} 并没有利用，而这些最新计算出来的分量可能比旧的 u_h^k 分量更好地近似于精确解分量 u_1，…，u_{i-1}。GS 方法对这些最新分量 $u_j^{k+1}(j=1, 2, \cdots, i-1)$ 作了合理利用。

若网格节点采用整体累加规则编号，固定 i，则解离散方程(2.1)的 GS 方法计算公式为

$$u_i^{k+1}=\frac{1}{a_{ii}}\Big(f_i-\sum_{j=1}^{i-1}a_{ij}u_j^{k+1}-\sum_{j=i+1}^{N}a_{ij}u_j^k\Big).$$

若考虑到离散格式和 GS 方法的局部特征，则上式可进一步简化，此时 GS 方法的格式点集分解情形为

$$V^+=\{j\,|\,j\text{ 是 }i\text{ 的邻点且 }j<i\},\qquad V^0=\{i\},$$
$$V^-=\{j\,|\,j\text{ 是 }i\text{ 的邻点且 }j>i\};$$

于是 GS 迭代公式又可改写为

$$u_i^{k+1}=\frac{1}{a_{ii}}\Big(f_i-\sum_{j\in V^+}a_{ij}u_j^{k+1}-\sum_{j\in V^-}a_{ij}u_j^k\Big), \tag{2.13}$$

此式可用于非规则网格分划情形下离散方程的求解。

对直角三角形分划和矩形分划，网格节点可按坐标形式编号，在应用 9 点格式时 GS 方法的符号格式和格式点集分解情形为

$$V=\begin{pmatrix}- & - & -\\ + & 0 & -\\ + & + & +\end{pmatrix},\qquad \begin{aligned}&V^0=\{(0,\ 0)\},\\ &V^+=\{(k_1,\ k_2)\,|\,k_1+k_2<0\text{且}k_2<0\},\\ &V^-=\{(k_1,\ k_2)\,|\,k_1+k_2>0\text{且}k_2>0\}.\end{aligned}$$

于是利用离散格式和符号格式就很容易写出节点按坐标编号情况下的 GS 方法计算公式

$$u_{ij}^{k+1}=\frac{1}{s_{0,0}}(\bar{f}_{ij}-s_{-1,0}u_{i-1,j}^{k+1}-s_{-1,-1}u_{i-1,j-1}^{k+1}-s_{0,-1}u_{i,j-1}^{k+1}$$

$$-s_{1,-1}u_{i+1,j-1}^{k+1}-s_{-1,1}u_{i-1,j+1}^{k}-s_{0,1}u_{i,j+1}^{k}$$

$$-s_{1,1}u_{i+1,j+1}^{k}-s_{1,0}u_{i+1,j}). \qquad (2.14)$$

用此式计算，每个完整的 GS 迭代步至多需要 $9N$ 次乘(除)法和 $9N$ 次加(减)法运算；实算时只需 1 套存贮单元，存放 u_h^k。

利用局部 Fourier 分析方法，可分析 GS 方法的收敛性。例如对 5 点离散格式情形，运用定理 2.2 中的结论，有

$$\lambda^0=4, \qquad \lambda^+=-[\exp(-i\theta_1)+\exp(-i\theta_2)],$$

$$\lambda^-=-[\exp(i\theta_1)+\exp(i\theta_2)], \qquad A(\theta,h,1)=-\lambda^-/(\lambda^0+\lambda^+),$$

$$|A(\theta,h,1)|^2=\frac{(\cos\theta_1+\cos\theta_2)^2+(\sin\theta_1+\sin\theta_2)^2}{(4-\cos\theta_1-\cos\theta_2)^2+(\sin\theta_1+\sin\theta_2)^2};$$

用 $\theta_2=\theta_1=\pi h$ 和 $\cos\theta_1=(1-\theta_1^2)/2$ 代入上式，即得

$$\rho(T_h)=\max|A(\theta,\ h,\ 1)|\approx 1-\pi^2h^2=1-O(h^2).$$

用同样的办法可算得 9 点差分格式和双线性有限元格式的 GS 迭代矩阵谱半径都是 $1-O(h^2)$，GS 方法的收敛速度也都与 h^2 同阶。

GS 方法的矩阵表示式为

$$(L_h^l+L_h^d)u_h^{k+1}+L_h^r u_h^k=f_h,$$

或

$$\begin{aligned}u_h^{k+1}&=(L_h^l+L_h^d)^{-1}(f_h-L_h^r u_h^k)\\&=-(L_h^l+L_h^d)^{-1}L_h^r u_h^k+(L_h^l+L_h^d)^{-1}f_h\\&=[I-(L_h^l+{}_h^d)^{-1}L_h]u_h^k+(L_h^l+L_h^d)^{-1}f_h, \qquad (2.15)\end{aligned}$$

故 GS 法迭代矩阵为 $-(L_h^l+L_h^d)^{-1}L_h^r$ 或 $I-(L_h^l+L_h^d)^{-1}L_h$。

§5 SOR 方法

SOR 方法(Successive Over Relaxation Method)是在

GS 方法基础上的一种加速方法。该方法首先利用 GS 方法定义辅助量$\bar{u}_i^{k+1}$：

$$\bar{u}_i^{k+1}=\frac{1}{a_{ii}}\left(f_i-\sum_{j=1}^{i-1}a_{ij}u_j^{k+1}-\sum_{j=i+1}^{N}a_{ij}u_j^k\right);$$

再由 u_i^k 与 $\bar{u}_i^{k+1}$作加权平均：

$$u_i^{k+1}=(1-\omega)u_i^k+\omega\bar{u}_i^{k+1}=u_i^k+\omega(\bar{u}_i^{k+1}-u_i^k);$$

就得到 SOR 方法的计算公式

$$\begin{aligned}u_i^{k+1}&=u_i^k+\frac{\omega}{a_{ii}}\left(f_i-\sum_{j=1}^{i-1}a_{ij}u_j^{k+1}-\sum_{j=i}^{N}a_{ij}u_j^k\right)\\&=(1-\omega)u_i^k+\frac{\omega}{a_{ii}}\left(f_i-\sum_{j=1}^{i-1}a_{ij}u_j^{k+1}-\sum_{j=i+1}^{N}a_{ij}u_j^k\right),\end{aligned}$$

其中ω为松弛因子，显然$\omega=1$时，SOR 方法就是GS方法。

SOR 方法的格式点集分解同于 GS 方法的情形，由此可写出用于一般非规则网格分划离散方程的 SOR 方法计算公式：

$$u_i^{k+1}=(1-\omega)u_i^k+\frac{\omega}{a_{ii}}\left(f_i-\sum_{j\in V^+}a_{ij}u_j^{k+1}-\sum_{j\in V^-}a_{ij}u_j^k\right).\tag{2.16}$$

同样也可写出用于节点坐标式编号情况下的 9 点离散格式 SOR 方法计算公式

$$\begin{aligned}u_{i,j}^{k+1}=(1-\omega)u_{i,j}^k+\frac{\omega}{s_{0,0}}(\bar{f}_{ij}&-s_{-1,0}u_{i-1,j}^{k+1}-s_{-1,-1}u_{i-1,j-1}^{k+1}\\-s_{0,-1}u_{i,j-1}^{k+1}&-s_{1,-1}u_{i+1,j-1}^{k+1}-s_{-1,1}u_{i-1,j+1}^k-s_{0,1}u_{i,j+1}^k\\&-s_{1,1}u_{i+1,j+1}^k-s_{1,0}u_{i+1,j}^k).\end{aligned}\tag{2.17}$$

下面讨论 5 点离散格式 SOR 方法的收敛速度。参照(2.17)式很容易写出其迭代矩阵T_h，从而可写出其迭代矩阵T_h的特征值问题

$$T_hv=\lambda_\omega v$$

的分量形式：

$$(1-\omega)v_{ij}+\frac{\omega}{s_{0,0}}(s_{1,0}v_{i+1,j}+s_{0,1}v_{i,j+1})$$

$$=\lambda_{\omega}\left[v_{ij}-\frac{\omega}{s_{0,0}}(s_{-1,0}v_{i-1,j}+s_{0,-1}v_{i,j-1})\right].$$

令 $v_{ij}=\lambda_{\omega}^{(i+j)/2}\varphi_{ij}$，将其代入上式并消去因子 $\lambda_{\omega}^{(i+j)/2}$，得到

$$(1-\omega)\varphi_{ij}+\lambda_{\omega}^{1/2}\frac{\omega}{s_{0,0}}(s_{1,0}\varphi_{i+1,j}+s_{0,1}\varphi_{i,j+1})$$

$$=\lambda_{\omega}\left[\varphi_{ij}-\lambda_{\omega}^{-1/2}\frac{\omega}{s_{0,0}}(s_{-1,0}\varphi_{i-1,j}+s_{0,-1}\varphi_{i,j-1})\right], \qquad (2.18)$$

整理得

$$\frac{\lambda_{\omega}+\omega-1}{\omega\lambda_{\omega}^{1/2}}\varphi_{ij}=\frac{1}{s_{0,0}}(s_{-1,0}\varphi_{i-1,j}+s_{0,-1}\varphi_{i,j-1}$$

$$+s_{1,0}\varphi_{i+1,j}+s_{0,1}\varphi_{i,j+1}).$$

因为 v_{ij} 不恒为零，因此 φ_{ij} 也不恒为零，上式正表明

$$\mu=\frac{\lambda_{\omega}+\omega-1}{\omega\lambda_{\omega}^{1/2}} \qquad (2.19)$$

恰为解5点离散格式的 Jacobi 方法迭代矩阵的特征值。

反之，若 μ 是5点离散格式 Jacobi 迭代的特征值，且 (2.19)式成立，则 λ_{ω} 必为5点离散格式 SOR 方法的特征值。事实上，当 $\lambda_{\omega}\neq 0$ 时，有不恒为零的 φ_{ij} 使(2.19)式成立，再令$\varphi_{ij}=\lambda_{\omega}^{(-i-j)/2}v_{ij}$，经过简单计算就得到(2.18)式，即 λ_{ω} 为5点离散格式 SOR 迭代矩阵的特征值；若 $\lambda_{\omega}=0$，从(2.19) 式的变形式得 $\omega=1$，此时也容易证得0确为5点离散格式 SOR 方法的一个特征值。

假定 μ 是实数，(2.19)式可视为 $\lambda_{\omega}^{1/2}$的实系数二次多项式，从而可知当$0<\omega<2$ 且 Jacobi 迭代收敛时，SOR 方法收敛。

求5点离散格式 SOR 方法的最佳松弛因子，相当于求 ω 使

(2.19)式的根的最大值达到最小。解(2.19)式，有

$$\lambda_\omega^{1/2}=\frac{\omega\mu}{2}\pm\sqrt{\frac{\omega^2\mu^2}{4}-(\omega-1)}.$$

当$(\omega-1)>\omega^2\mu^2/4$时，(2.19)式只有复根，其模为$(\omega-1)^{1/2}$，它随$\omega$增加而增加；当$(\omega-1)<\omega^2\mu/4$时，(2.19)式有相异二实根，其最大模为

$$\max|\lambda_\omega^{1/2}|=\frac{\omega|\mu|}{2}+\sqrt{\frac{\omega^2\mu^2}{4}-\omega+1}<1,$$

而

$$\frac{d}{d\omega}(\max|\lambda_\omega^{1/2}|)=\frac{|\mu|\max|\lambda_\omega^{1/2}|-1}{2\left(\max|\lambda_\omega^{1/2}|-\frac{1}{2}\omega|\mu|\right)}<0$$

表明$\max|\lambda_\omega^{1/2}|$为$\omega$的单调下降函数。综合上述两种情形，当$\omega-1=\omega^2\mu^2/4$时，即当

$$\omega=2/(1+\sqrt{1-\mu^2})$$

时，$\max|\lambda_\omega^{1/2}|$达到极小，此时

$$\max|\lambda_\omega|=\omega-1=\frac{1-\sqrt{1-\mu^2}}{1+\sqrt{1-\mu^2}}. \tag{2.20}$$

注意，μ是Jacobi迭代的任意特征值，上式中$\max|\lambda_\omega|$是μ的上升函数，故知最佳松弛因子应取为

$$\omega_{opt}=\frac{2}{1+\sqrt{1-\mu_{max}^2}}. \tag{2.21}$$

由于松弛因子$\omega_{opt}>1$，故称为**超松弛**。

图2.1画出了$\max|\lambda_\omega|$作为ω的函数图形。

从上图看到，当$\omega<\omega_{opt}$时，虽然ω偏离ω_{opt}很小，但$\max|\lambda_\omega|$却变化很大；而当$\omega>\omega_{opt}$时，它是ω的线性函数且斜率为1，

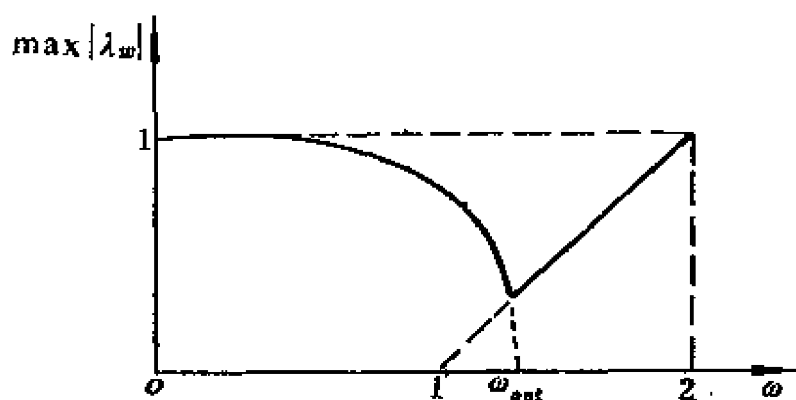


图2.1

变化缓慢。因此实际计算时，近似的最佳松弛因子应比 ω_{opt}略大一点较为保险。

现在估计模型问题SOR 方法的收敛速度。由本节结论，模型问题 Jacobi 迭代最大特征值是

$$\mu_{max} \approx 1 - \frac{\pi^2 h^2}{2},$$

将其代入(2.20)式，得到

$$\max|\lambda_\omega| \approx \frac{1-\sqrt{1-\left(1-\frac{\pi^2 h^2}{2}\right)^2}}{1+\sqrt{1-\left(1-\frac{\pi^2 h^2}{2}\right)^2}} \approx 1 - 2\pi h = 1 - O(h),$$

从而 SOR 方法的收敛速度是

$$R(T_h(\omega_{opt})) = -\ln|\lambda_{\omega_{opt}}| \approx 2\pi h.$$

总之，在每个松弛迭代步中，SOR 方法的计算量与 Jacobi 方法和 GS 方法的计算量相当，实算时它只需一组存贮单元即可工作；若选择好最佳松弛因子，SOR 方法的收敛速度与 h 同阶，比 Jacobi 方法和 GS 方法的收敛速度改进了 h 的一个阶次。因此 SOR 方法至今仍是解离散微分方程的主要方法之一。

值得指出的是，上述推导只是针对 5 点离散格式进行的，如何对各种离散格式建立一般的松弛理论至今未能见到有关的文章。

在实际应用中，ω_{opt}是不好准确选取的，这里介绍一种近似估计方法。先用 GS 方法计算若干步，根据计算特征值的**乘幂法**，可近似算出 GS方法的最大特征值 λ_1

$$\lambda_1 \approx \sum_{i,j} |u_{ij}^{k+1}| / \sum_{i,j} |u_{ij}^{k}|;$$

再根据 5 点格式情形的 GS 方法特征值 λ_1 和 Jacobi 方法特征值 μ 的关系 $\lambda_1 = \mu^2$，且利用(2.21)式，即可取得

$$\omega_{opt} \approx \frac{2}{1 + \sqrt{1 - \lambda_1}}.$$

作为本节结束和为了以后使用方便，这里给出解离散方程(2.1)的 SOR 方法的矩阵表示式

$$u_h^{k+1} = (L_h^d + \omega L_h^l)^{-1}[(1-\omega)L_h^d - \omega L_h^u]u_h^k + \omega(L_h^d + \omega L_h^l)^{-1} f_h,$$

这里假定$(L_h^d + L_h^l)^{-1}$存在，由此可知 SOR 方法的迭代矩阵为

$$(L_h^d + \omega L_h^l)^{-1}[(1-\omega)L_h^d - \omega L_h^u] \text{或} I - \omega(L_h^d + L_h^l)^{-1} L_h.$$

§6 逐线松弛法

6.1 逐线 Jacobi 方法

设矩形定解区域采用矩形或直角三角形网格分划，节点采用坐标式编号(i, j)，i、$j = 1, \cdots, n$，$N = n \times n$；令$U_i = (u_{i1}, u_{i2}, \cdots, u_{in})^T$，$F_i = (f_{i1}, f_{i2}, \cdots, f_{in})^T$；则用离散格式(2.5)表示的离散方程可改写成按线(即按行指标 i 或按列指标 j) 分块矩阵的表示式

$$A_i U_{i-1} + B_i U_i + C_i U_{i+1} = F_i, \quad i = 1, \cdots, n, \qquad (2.22a)$$

它的逐线 Jacobi 迭代(记为 LJacobi)公式是

$$A_i U_{i-1}^k + B_i U_i^{k+1} + C_i U_{i+1}^k = F_i, \quad i = 1, \cdots, n. \qquad (2.22b)$$

其计算过程是这样的：求解一个三对角方程组，获得一个列向量 U_i^{k+1}；解 n 个三对角方程组，才完成 LJacobi 的一个完整迭代步；实算中需用 2 套存贮单元。

将(2.22b)式中的已知量合并为方程组右端项，各需 $6N$次乘(除)运算和加(减)运算；用追赶法解一个 3对角方程组，约需$5N$次乘(除)和 $5N$次加减运算；故 LJacobi 一个完整迭代步共需要 $11N$ 次乘(除)和 $11N$ 次加(减)运算。

LJacobi 迭代公式(2.22b)可表示为

$$[S_k^0]U_i^{k+1}(x)+[S_k^-]U_i^k(x)=F_i(x), \tag{2.23}$$

$$x\in\{(x_i,\ y_j)\mid j=1,\ \cdots,\ n\},\qquad i=1,\ \cdots,\ n,$$

其中，$[S_k]=[S_k^+]+[S_k^0]+[S_k^-]$，

$$[S_k^+]=\varnothing,\ [S_k^0]=\begin{pmatrix}0 & s_{0,1} & 0\\ 0 & s_{0,0} & 0\\ 0 & s_{0,-1} & 0\end{pmatrix},\qquad [S_k^-]=\begin{pmatrix}s_{-1,1} & 0 & s_{1,1}\\ s_{-1,0} & 0 & s_{1,0}\\ s_{-1,-1} & 0 & s_{1,-1}\end{pmatrix}.$$

(2.23)式的分量形式是

$$\begin{aligned}&s_{0,1}u_{i,j+1}^{k+1}+s_{0,0}u_{i,j}^{k+1}+s_{0,-1}u_{i,j-1}^{k+1}\\ &=f_{i,j}-s_{-1,1}u_{i-1,j+1}^k-s_{-1,0}u_{i-1,j}^k-s_{-1,-1}u_{i-1,j-1}^k\\ &-s_{1,1}u_{i+1,j+1}^k-s_{1,0}u_{i+1,j}^k-s_{1,-1}u_{i+1,j-1}^k,\quad j=1,\cdots,n.\end{aligned} \tag{2.24}$$

利用公式(2.23)、(2.24)和定理 2.2，可分析 LJacobi 方法的收敛特性。为此，记 LJacobi 迭代矩阵的特征值和谱半径为 μ 和$|\mu|_{max}$。

对于 5 点离散格式情形，有

$$V=\begin{pmatrix} & 0 & \\ - & 0 & -\\ & 0 & \end{pmatrix},\quad \begin{aligned}&\lambda^+=0,\qquad \lambda^-=-\exp(i\theta_1)-\exp(-i\theta_1),\\ &\lambda^0=4-\exp(i\theta_2)-\exp(-i\theta_2),\end{aligned}$$

$$\mu=-\lambda^-/\lambda^0=\cos\theta_1/(2-\cos\theta_2).$$

利用无穷小代换，即用 $\theta_1=\theta_2$、$\cos\theta_1\approx1-\theta_1^2/2$、$\theta_1=\pi h$ 代入上式，得

$$|\mu|_{max}\approx1-\pi^2h^2=1-O(h^2). \tag{2.25}$$

所以，5 点离散格式的 LJacobi 迭代敛速比逐点 Jacobi 迭代敛速快 2 倍。

用同样办法可计算其它格式的 LJacobi 迭代敛速，例如正方形双线性元格式情形，有$|\mu|_{max}=1-\pi^2h^2$。

6.2 逐线SOR方法

解(2.22a)式的逐线 SOR 方法(记为 LSOR) 也是常用的有效方法之一，其迭代公式为

$$[S_k^0]U_i^{k+1}(x)=(1-\omega)[S_k^0]U_i^k(x)+\omega(F_i(x)-[S_k^+]U_i^{k+1}(x)-[S_k^-]U_i^k(x)),\quad x\in\{(x_i,\ y_j)\,|\,j=1,\cdots,n\},\ i=1,\cdots,n,\tag{2.26}$$

其中

$$[S_k^+]=\begin{pmatrix}s_{-1,1}&0&0\\s_{-1,0}&0&0\\s_{-1,-1}&0&0\end{pmatrix},\quad [S_k^0]=\begin{pmatrix}0&s_{0,1}&0\\0&s_{0,0}&0\\0&s_{0,-1}&0\end{pmatrix},$$

$$[S_k^-]=\begin{pmatrix}0&0&s_{1,1}\\0&0&s_{1,0}\\0&0&s_{1,-1}\end{pmatrix};$$

分量形式为

$$\begin{cases}u_{i,j}^{k+1}=u_{i,j}^k+\omega(\tilde{u}_{i,j}^{k+1}-u_{i,j}^k),\\ s_{0,1}\tilde{u}_{i,j+1}^{k+1}+s_{0,0}\tilde{u}_{i,j}^{k+1}+s_{0,-1}\tilde{u}_{i,j-1}^{k+1}=f_{i,j}\\ -s_{-1,1}u_{i-1,j+1}^{k+1}-s_{-1,0}u_{i-1,j}^{k+1}-s_{-1,-1}u_{i-1,j-1}^{k+1}\\ -s_{1,1}u_{i+1,j+1}^k-s_{1,0}u_{i+1,j}^k-s_{1,-1}u_{i+1,j-1}^k,\\ j=1,\ 2,\ \cdots,\ n.\end{cases}\tag{2.27}$$

如同上节的分析一样，一个完整的 LSOR 迭代步需要解n个 3 对角方程组；对于 9 点离散格式所确定的离散方程，一个完整的 LSOR 迭代步需要 11N 次乘(除)和 11N 次加(减) 运算；LSOR 方法在实算中只要一套存贮单元。

由逐点 SOR 方法的收敛性分析不难知道：若(2.26) 式的系数矩阵对称正定且 LJacobi 方法收敛，则 LSOR 方法也收敛；LSOR 方法的特征值λ_ω、最佳松弛因子ω_{opt}和 LJacobi 方法的

特征值 μ、谱半径$|\mu|_{max}$适合下述关系：

$$\begin{cases} \mu = \dfrac{\lambda_\omega + \omega - 1}{\omega \lambda_\omega}, \\ \omega = \dfrac{2}{1+\sqrt{1-\mu_{max}^2}}, \\ \max|\lambda_{\omega_{opt}}| = \omega_{opt} - 1 \end{cases} \tag{2.28}$$

利用(2.28)式和 LJacobi 方法的特征值 μ，不难求出LSOR 方法的敛速。对于 5 点离散格式，由(2.25)式和(2.28)式有

$$\max|\lambda_{\omega_{opt}}| = \frac{1-\sqrt{1-\mu_{max}^2}}{1+\sqrt{1-\mu_{max}^2}} = \frac{1-\sqrt{1-\left(\dfrac{\cos\pi h}{2-\cos\pi h}\right)^2}}{1+\sqrt{1-\left(\dfrac{\cos\pi h}{2-\cos\pi h}\right)^2}}$$

$$\approx 1-2\sqrt{2}\pi h,$$

$$\omega_{opt} = \frac{2}{1+\sqrt{1-\mu_{max}^2}} \approx 2-2\sqrt{2}\pi h.$$

故 5 点离散格式的 LSOR 方法的最佳敛速比逐点 SOR 方法的最佳敛速快$\sqrt{2}$倍。

§7 RB松弛法

设对矩形区域作国际象棋棋盘式分划，如图 2.2 所示，还可参见图 4.1，其中“∘”点和“*”点分别表示“**红点**”和“**黑点**”，**棋盘式区域分划简记为RB分划**。节点编号仍采用坐标编号规则。

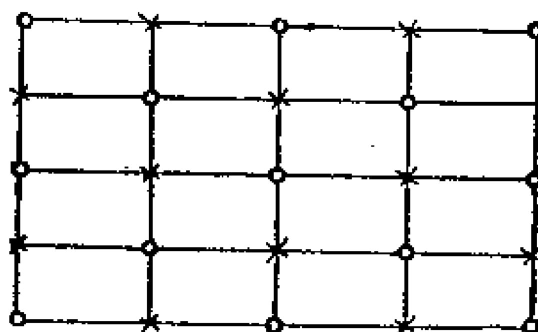


图2.2

RB 松弛法由 2 个部分松弛步组成：第一个部分步中，只对“黑点”(或“红点”)作某种松弛迭代，直到所有

的“黑点”都作完为止；第二个部分步中，只对“红点”(或“黑点”)作某种松弛迭代，直到全部“红点”作完为止。2个松弛步完成后即完成对所有网格节点施行了一个完整的RB松弛迭代步。在RB松弛法中，“红点”集合和“黑点”集合可分别表示为

$$\text{红点：}\{x\in\Omega_h \mid x=kh,\ k_1+k_2=\text{偶数}\},$$

$$\text{黑点：}\{x\in\Omega_h \mid x=kh,\ k_1+k_2=\text{奇数}\};$$

松弛方法可采用 ω-Jacobi 迭代和 GS 迭代等。

例如，对Poisson问题使用9点离散格式$[S_h]$，此时，RB-GS迭代计算公式为

step 1

$$\begin{aligned} u_{ij}^{k+1}=\frac{1}{s_{0,0}}(f_{ij}&-s_{-1,-1}u_{i-1,j-1}^{k+1}-s_{0,-1}u_{i,j-1}^{k+1}\\ &-s_{1,-1}u_{i+1,j-1}^{k+1}-s_{-1,0}u_{i-1,j}^{k+1}-s_{0,1}u_{i,j+1}^{k}\\ &-s_{-1,1}u_{i-1,j+1}^{k}-s_{1,1}u_{i+1,j+1}^{k}-s_{1,0}u_{i+1,j}^{k}),\end{aligned}$$

当 $i+j=$ 偶数时，

$$u_{ij}^{k+1}=u_{ij}^{k},\qquad \text{当 } i+j=\text{奇数时；}$$

step 2

将 step 1 中的“红”“黑”点对调，仍按 step 1 中的计算公式计算。

对5点离散格式情形，读者不难看出，一个完整的RB-Jacobi 迭代步相当于2步逐点 Jacobi 迭代，两者的计算工作量相当，RB-Jacobi 迭代的收敛因子为逐点 Jacobi 迭代收敛因子的平方；一个完整的 RB-GS 迭代步相当于1步 Jacobi 迭代和一步 GS 迭代的效果。

RB 松弛法只需一套存贮单元，其计算量与逐点松弛法的相同，但敛速却提高了，因此 RB 松弛法是实用可取的。

§8 共轭梯度加速法

50年代初期，Hestenes 和 Stiefel[5]提出了一种解线代数方程组的新迭代法，叫作**“共轭梯度法”**，记为 CG 方法。CG 法虽是一个迭代法，但当不计舍人误差时，它以有限次迭代收敛于方程组的精确解。由于这和其它许多有趣的性质，CG法刚出现就引起了数值分析界的极大注意。但是，由于种种原因，这种方法曾消声匿迹了许多年[6,7]。从60年代中期以来，对 CG 方法的兴趣有一股较大的回潮。现在，人们越来越深刻地认识了 CG 方法的本质并扩大了它的应用。

8.1 最速下降法

设有线性离散方程组 $L_h u_h = f_h$，其中 u_h 和 f_h 都是N阶列向量，L_h 是 $N\times N$ 阶对称正定矩阵，为了导出并认识最速下降法，先注意如下两个事实。

第一，$L_h u_h = f_h$ 的解就是二次泛函

$$J(u_h)=\frac{1}{2}(L_h u_h,\ u_h)-(f_h,\ u_h)\quad (f_h、u_h\in R^N)$$

的极小解。事实上，若令 $J(u_h)$在 $\bar{u}_h$ 处取得极小值$J(\bar{u}_h)$，并设 $\forall v_h\in R^N$，$\lambda\in R$，则 $J(u_h)=J(\bar{u}_h+\lambda v_h)$可看作关于实参数 λ 的函数。再利用矩阵 L_h 的对称性和线性：

$$(L_h u_h,\ v_h)=(u_h,\ L_h v_h),$$

$$(L_h(\bar{u}_h+\overset{*}{u}_h),\ v_h)=(L_h\bar{u}_h,\ v_h)+(L_h\overset{*}{u},v_h),$$

即可将 $J(u_h)$改写为

$$\begin{aligned}\varphi(\lambda)&=J(u_h)=J(\bar{u}_h+\lambda v_h)\\&=J(\bar{u}_h)+\lambda(L_h\bar{u}_h-f_h,\ v_h)+\frac{\lambda^2}{2}(L_h v_h,\ v_h).\end{aligned}\qquad(2.29)$$

据经典微分法知，$J(u_h)$取得极小值必须满足条件：

$$\varphi'(\lambda)|_{\lambda=0}=0,\ \text{即}L_h\bar{u}_h=f_h;$$

$$\varphi''(\lambda)|_{\lambda=0}>0, \text{即}(L_h v_h, v_h)>0\text{或}L_h\text{正定。}$$

从而说明了第一个性质。

第二，$J(u_h)$的梯度方向是泛函$J(u_h)$在u_h处具有最大瞬时变化率的方向，它的负方向与线性离散方程组的残差向量一致。事实上，

$$\begin{aligned}\operatorname{grad}J(u_h) &= J'(u_h)=\left[\frac{1}{2}(L_h u_h, u_h)-(f_h, u_h)\right]' \\ &= \frac{1}{2}[(L_h u_h, I_h)+(L_h I_h, u_h)]-(f_h, I_h) \\ &= (L_h u_h-f_h, I_h)=L_h u_h-f_h,\end{aligned}$$

若u_h^k是u_h的近似值，则有

$$-\operatorname{grad}J(u_h^k)=f_h-L_h u_h^k=r_h^k,$$

其中$I_h\in R^N$是单位向量，r_h^k表示残差向量。

基于以上两点事实，即可方便形象地导出最速下降法。若设u_h^k是$L_h u_h=f_h$的某个近似解，则最速下降法通过下述途径改善近似值：在$\operatorname{grad}J(u_h^k)$方向上移动$u_h^k$，即$u_h^{k+1}=u_h^k+\lambda_k\operatorname{grad}J(u_h^k)$；再选择$\lambda_k$使$J(u_h^{k+1})$取得极小值，$\lambda_k$可利用(2.29)式确定；最速下降法计算公式是：

$$\begin{cases}u_h^{k+1}=u_h^k+\lambda_k r_h^k, \quad k=0, 1, \cdots, \quad u_h^0\text{任意}, \\ r_h^k=f_h-L_h u_h^k \\ \lambda_k=(r_h^k, r_h^k)/(L_h r_h^k, r_h^k).\end{cases}$$

当L_h病态(参见 §10)时，最速下降法敛速可能很慢[8]。然而，只要改变一下方向的选择办法，就可得到效果大不相同的"共轭梯度法"。

8.2 共轭梯度法

设向量p_h^i和$p_h^j\in R^N$，若$(p_h^i, L_h p_h^j)=0$，$i\neq j$，则称p_h^i和p_h^j是**A-共轭**或**A-正交**的。当然，N个A-正交的向量$\{p_h^i, i=0, 1, \cdots, N-1\}$可作$N$维空间的基底。

CG 法通过下述途径改善了u_h^k：在p_h^k的方向上移动u_h^k，有

$u_h^{k+1}=u_h^k+\lambda_k p_h^k$; 再选择 λ_k 使 $J(u_h^{k+1})$ 取得极小值; p_h^k 和 p_h^{k-1} 要是 A-共轭的，即 $(p_h^k,\ L_h p_h^{k-1})=0$。这样就不难确定和理解 CG 方法的计算公式：

$$
\begin{cases}
u_h^{k+1}=u_h^0+\lambda_k p_h^k, \quad k=0,\ 1,\ \cdots, \quad u_h^0\text{任意},\\
p_h^k=r_h^k+\alpha_k p_h^{k-1},\ k=1,\ 2,\ \cdots,\ p_h^0=r_h^0,\\
\alpha_k=-(r_h^k,\ L_h p_h^{k-1})/(p_h^{k-1},\ L_h p_h^{k-1}),\ k=1,2,\cdots,\\
\lambda_k=(p_h^k,\ r_h^k)/(p_h^k,\ L_h p_h^k),\ k=0,\ 1,\ \cdots,\\
r_h^k=f_h-L_h u_h^k,\ k=0,\ 1,\ \cdots。
\end{cases}
$$

可以证明，上式中 α_k、λ_k 和 r_h^k 等价于

$$
\alpha_k=(r_h^k,\ r_h^k)/(r_h^{k-1},\ r_h^{k-1}),
$$

$$
\lambda_k=(r_h^k,\ r_h^k)/(p_h^k,\ L_h p_h^k),
$$

$$
r_h^k=r_h^{k-1}-\lambda_{k-1}L_h p_h^{k-1},
$$

所以较为方便实用的CG法计算公式为

$$
\begin{cases}
u_h^{k+1}=u_h^k+\lambda_k p_h^k,\ k=0,\ 1,\ \cdots,\ u_h^0\text{任意},\\
p_h^k=r_h^k+\alpha_k p_h^{k-1},\ k=1,\ 2,\ \cdots,\ p_h^0=r_h^0,\\
\alpha_k=(r_h^k,\ r_h^k)/(r_h^{k-1},\ r_h^{k-1}),\ k=1,2,\cdots,\\
\lambda_k=(r_h^k,\ r_h^k)/(p_h^k,\ L_h r_h^k),\ k=0,1,\cdots,\\
r_h^k=f_h-L_h u_h^k, \qquad k=0,1,\cdots
\end{cases} \tag{2.30}
$$

由(2.30)式可知，使用CG法需同时存放三个向量 u_h^k、p_h^k 和 r_h^k。CG法实际上同时进行了两个迭代过程，一是 p_h^k 的 A-正交化过程，再是近似解极小化序列构造过程。若采用 9 点离散格式，CG 法中计算一次内积(如 $(p_h^k,\ r_h^k)$ 等)需 N 次乘(除)和 N 次加(减)运算；计算 $L_h p_h^k$ 需 $9N$ 次乘(除)和 $9N$ 次加(减)运算；因此由(2.30)式容易估计出一个 CG 迭代步约需 $24N$ 次乘(除) 和 $24N$ 次加(减)运算。

[5]中还证明了残差向量和方向向量之关系

$$
(r_h^i,\ r_h^j)=0,\ (p_h^i,\ L_h p_h^j)=0,\ i\neq j,
$$

$$
(r_h^i,\ L_h p_h^j)=0,\ i\neq j,\ i\neq j+1,
$$

可见 r_h^i 是相互正交的，而 p_h^i 是相互 A-正交的。

利用共轭梯度法，理论上不超过N步即可准确地求得u_h，所以可称CG法是N步准确解法。

事实上由于p_h^0，…，p_h^{N-1}是相互A-正交的，所以

$$u_h - u_h^0 = \sum_{i=0}^{N-1} c_i p_h^i, \quad (u_h - u_h^0, \ L_h p_h^j) = c_j (p_h^j, \ L_h p_h^j),$$

$$c_j = (u_h - u_h^0, \ L_h p_h^j)/(p_h^j, \ L_h p_h^j), \quad j = 0, \ 1, \ \cdots, \ N-1.$$

利用归纳法可证$c_k = \lambda_k$，$k = 0$，1，…，$N-1$。最后再利用u_h^k的递推定义，就有

$$u_h^N = u_h^0 + \sum_{i=0}^{N-1} \lambda_i p_h^i = u_h^0 + \sum_{i=1}^{N-1} c_i p_h^i = u_h.$$

8.3 CG法的几何意义

因为$J(u_h^k)$=常数是一个超椭球，其中心正是$L_h u_h = f_h$的解，其表面的负梯度方向（负法向）与残差向量r_h^k具有共同的方向，它指向**椭球中心**u_h。这表明，从某一个u_h^0出发，沿着r_h^0方向作直线移动，能使$J(u_h^0)$达到最大瞬时下降。例如按(2.30)式选取λ_0，在$u_h^0 + \lambda_0 r_h^0$处泛函J达到极小。这就是数值方法中广泛应用的**梯度法**的一个例子

然而，若只沿梯度方向修正近似解，则不能保证N步内求出准确解。为此，CG法中采用了线性独立的共轭向量列$\{p_h^j\}$作为方向向量。可以证明[9]，CG法中的方向向量p_h^j是梯度向量$r_h^j = \mathrm{grad} J(u_h^j)$在$p_h^j$，…，$p_h^{N-1}$张成的线性空间中投影的常数倍，共轭梯度法这一名称正是源于这一事实和p_h^j的A-共轭性。

当u_h^0沿p_h^0方向移动到$u_h^1 = u_h^0 + \lambda_0 p_h^0$时，求$\lambda_0$的极小化过程表明$u_h^1$是在由$p_h^0$，…，$p_h^{N-1}$张成的$N$维空间中的投影；当$u_h^1$沿$p_h^1$方向移动到$u_h^2 = u_h^1 + \lambda_1 p_h^1$时，由于$p_h^0$和$p_h^1$的$A$正交性，求$\lambda_1$的极小化过程则表明$u_h^2$是在由$p_h^1$，…，$p_h^{N-1}$这个$N-1$维空间中的投影，如此下去，由$\{p_h^j\}$的正交性可知，计算过程理论上需要$N$步就可求出线性离散方程组的解$u_h$。

8.4 CG法的优缺点

CG 法引起人们重视的原因不仅在于它是一种精确解法，它可用于更广泛类型的线性方程组求解，该方法中的一些概念还可能有更广泛的应用；而且它具有迭代法的特点，敛速$>2\pi h$(就模型问题)。作为迭代法，它虽要用三套存贮单元，但比消去法所用的存贮量少得多；它可用$\|r_h^k\|$作为计算控制量，若N步迭代达不到精度要求，它允许将u_h^N作为初始近似值重新开始迭代，直到$\|r_h^k\|$足够小。

CG 法的缺点在舍入误差的积累方面。该方法解大型稀疏线性方程组所用的总计算量与消元法差不多，但由于乘除运算次数多，舍人误差影响大，p_h^k很快丧失正交性，所以不仅几乎不可能在N步迭代内求出方程组的解，而且误差积累也会影响解的精度。经验表明，用 CG 法求得的解的精度比消去法要稍微差一点，但人们至今未能给出最后近似解的误差界限。因此，如何改进 CG 法，使其对误差积累不敏感，改进 CG 法的数值解精度是人们共同关心的问题。

§9 迭代方法的比较

迭代法和以消去法为基础的直接法比较，迭代法的共同优点是计算程序简单且存贮量少。例如，就存贮量而言，若用直接法求解大型稀疏的离散微分方程，无论采用半带宽存贮还是变带宽存贮，都要存贮大量的零元素，存贮量大且与网格节点编号密切相关；若使用松弛迭代法，则只需存贮大型稀疏矩阵的非零元素，存贮量不受节点编号的影响，存贮量达最佳水平。特别是在坐标式编号的一致网格剖分情况下，采用局部离散格式，不需要存贮系数阵信息就可使用迭代法求解。因此，评判和比较各种迭代法的标准可局限于其自身的性质，主要依据其收敛速度，每个迭代步所需的计算量和存贮量这三个指标。现将本章中所述的几种基

本迭代法的这些指标列于表 2.1 中。由于每迭代步所需乘(除)和加(减)的运算次数大体相当，所以表中只填写了乘(除)运算次数，并用“(×)”号表示，这些数值是按 9 点离散格式情形估算的；另外，表中有关敛速的数值，是模型问题，按 5 点离散格式计算的。

表2.1

迭代方法	每步迭代需存向量数	每步迭代所需运算量	模型问题的敛速
ω-Jacobi	2	$10N(\times)$	$\omega\pi^2h^2/2$
GS	1	$9N(\times)$	π^2h^2
SOR	1	$10N(\times)$	$2\pi h$
LJacobi	2	$11N(\times)$	π^2h^2
LSOR	1	$11N(\times)$	$2\sqrt{2}\pi h$
RB-Jacobi	2	$10N(\times)$	π^2h^2
RB-GS	1	$9N(\times)$	$3\pi^2h^2/2$
CG	8	$24N(\times)$	$>2\pi h$

从迭代法所能应用的矩阵类型方面看，LSOR 方法和共轭梯度法较好，它们的适应范围较广；从收敛速度方面看，共轭梯度法、SOR 方法、RB-SOR 和 LSOR 方法较好；从每迭代步的计算量方面看，SOR 和 RB-SOR 方法较好；从所需存贮单元多少方面看，SOR、RB-SOR 和 LSOR 方法较好。对于一般具有正定对称阵的离散方程来说，RB-SOR 和 SOR 方法可以取得比较全面的优势。

除具有精确解法性质的共轭梯度法外，迭代法的敛速都与网格步长 h 有关，对于敛速为 $O(h^2)$的迭代法，其迭代矩阵的谱半径为 $1-O(h^2)$，这不仅表明要减少初始误差 $1/e$ 倍需要 $O(N^2)$ 次

迭代步，误差越小收敛越慢，而且表明当网格步长 $h\to 0$ 时，迭代矩阵的性质变坏，敛速更差，因此，如何改进或构造新的迭代方法，使其具有快敛速、少工作，甚至达到敛速与 h 无关和计算量为 $O(N)$ 这个高标准，乃是极有意义的，也是人们关注和盼望的。

§10 迭代控制和迭代组合

10.1 迭代收敛控制及迭代解精度的改进

用某一种迭代法（迭代矩阵为 T_h）数值求解离散微分方程(2.1)，一般并不知道第 k 次迭代的近似值 u_h^k 和离散微分方程(2.1)准确解 u_h 的误差 $e_h^k=u_h^k-u_h$，但相邻两次迭代值之误差增量 $q_h^{k+1}=u_h^{k+1}-u_h^k$ 和残差向量 $d_h^k=f_h-L_hu_h^k$ 都是知道的，它们满足下述关系：

$$q_h^{k+1}=e_h^{k+1}-e_h^k=u_h^{k+1}-u_h^k,\qquad q_h^{k+1}=T_hq_h^k,$$

$$d_h^k=-L_he_h^k,\qquad d_h^{k+1}=L_hT_hL_h^{-1}d_h^k.$$

所以，在实际迭代过程中，可采用

$$|q_h^k|_\infty=\max_i|q_i^k|\leqslant\varepsilon_1,$$

或

$$|d_h^k|_\infty=\max_i|d_i^k|\leqslant\varepsilon_2,$$

作为迭代收敛的判据和控制，其中 q_i^k 和 d_i^k 分别是 q_h^k 和 d_h^k 的分量，ε_1 和 ε_2 是预先给定的小的正数。ε_1 或 ε_2 给得过宽，数值解精度不够；给得过严，会严重浪费计算工作量。

现以 5 点离散格式数值求解Poisson 问题为例，讨论用误差增量和残差向量控制迭代收敛之关系。

因 $f_h=h^2f_{ij}$，$\|L_h\|=O(h^2)$，$\|u-u_h\|$ 应保证至少有 2 阶精度，所以残差模量应取为 $\|d_h^k\|=\|f_h-L_hu_h^k\|=\|L_h\|\|u-u_h^k\|=O(h^4)$ 作为迭代收敛的控制量，再低就没有意义了。

当用误差增量的模作为迭代收敛控制量时，由于

$$q_h^k = u_h^k - u_h^{k-1} = e_h^k - e_h^{k-1} = (T_h - I)e_h^{k-1} = (T_h - I)(u_h^k - u_h)$$
$$= (T_h - I)[L_h^{-1}(f_h - d_h^{k-1}) - L_h^{-1}f_h] = (I - T_h)L_h^{-1}\, d_h^{k-1},$$
且考虑到L_h是常数矩阵时有 $L_h = T_h - I$，故知当L_h是常数矩阵时，应取$\|q_h^k\| = \|L_h\|\|L_h^{-1}\|O(h^4)$或简单地取$\|q_h^k\| = O(h^4)$作为迭代收敛控制量。

上述两种迭代控制量都可保证$\|u - u_h\|$至少有 $O(h^2)$ 量级的精度，在网格节点处，数值解精度要稍好一些。当然，在解曲面变化平缓，解的光滑性好且 L_h 是常数矩阵的情况下，取 $\|q_h^k\| = h^2 \sim h^3$作控制量也会有满意的效果。

用迭代法求解离散微分方程时，得到的是 u_h^k 而不是 u_h，u_h^k 的精度好坏主要受到机器舍入误差的影响和受到 L_h **条件数** $\mathrm{Cond}(L_h) = \|L_h^{-1}\|\|L_h\|$的影响。

由于机器舍入误差的存在，迭代近似解 u_h^k逼近 u_h 到一定程度后，数值解精度不再改进。若考虑到$\mathrm{Cond}(L_h)$并不坏，则人们可先用单字长计算(例如用$\|d_h^k\| \leqslant h^4$作为迭代控制量)，在获得某个解 u_h^k之后，再用双字长解误差校正方程：

$$L_h v_h = d_h^k = f_h - L_h u_h^k,$$

且用 $\bar{u}_h = u_h^k + v_h$ 作为最后的近似解，这样 $\bar{u}_h$ 的精度要比u_h^k好得多。

L_h 的条件数对离散方程(2.1)迭代近似解精度的影响是很大的，众所周知，迭代近似解 u_h^k和(2.1)式准确解之相对误差适合关系

$$\frac{\|u_h - u_h^k\|}{\|u_h\|} \leqslant \mathrm{Cond}(L_h)\,\frac{\|d_h^k\|}{\|f_h\|}.$$

病态(即 $\mathrm{Cond}(L_h) \gg 1$)离散方程是常见的，例如对于各向异性的椭圆型方程(参见(3.9) 式中 $a \ll b$或 $b \ll a$ 的情形)，其离散方程(参见(3.12)式)的系数矩阵常常是病态的。对病态离散方程，如不加选择地用迭代法求解，就会造成迭代近似解误差太大或严重失真。为此，可采用下述办法求解：

1° 采用某个方向的逐线松弛法求解；

2° 采用高精度的算术运算(如双字长运算)以便改善和减轻矩阵病态的影响；

3° 对方程组作预处理，即适当选择非奇异对角矩阵 D_h 和 C_h，使求解 $L_h u_h = f_h$ 的问题转化为求解等价方程组 $D_h^{-1} L_h C_h^{-1} (C_h u_h) = D_h^{-1} f_h$，在 $D_h^{-1} L_h C_h^{-1}$ 的条件数得到改善的情况下解得 $C_h u_h$，从而解出 u_h。

10.2 迭代过程的初期收敛性和迭代方法的组合应用

大家知道，在迭代过程中近似解误差满足关系式

$$\|e_h^k\| \leqslant \|T_h^k\| \|e_h^0\|.$$

若迭代矩阵 T_h 是对称阵时，其 Euclid 模与谱半径相同，即 $\rho(T_h) = \|T_h\|$；若 T_h 是非对称阵，则其欧代模与谱半径不相等，但满足

$$\rho(T_h) \leqslant \|T_h\| \leqslant N\rho(T_h),$$

其中，N 是近似解向量的维数。

当 T_h 是对称阵时，在 Jacobi 迭代方法中 $\|T_h\| = \rho(T_h)$，$\|T_h^k\| = \rho(T_h^k) = \rho^k(T_h)$，所以在 $\rho(T_h) < 1$ 时，近似解误差在 Jacobi 迭代过程中是按欧氏模单调下降的。

GS 方法中的迭代矩阵 T_h 不对称，但可以证明[8]

$$\|T_h^k\| = \sqrt{1+\mu^2} \cdot \mu^{2k-1}, \quad k = 1, 2, \cdots,$$

其中，μ 是逐点 Jacobi 迭代矩阵的谱半径，故近似解误差在 GS 迭代过程中也是按 Euclid 模单调下降的。

在带有最佳松弛因子的 SOR 方法中，迭代矩阵 $T_h(\omega_{opt})$ 也不对称，可以证明[8]。

$$\|T_h^k(\omega_{opt})\| = \left[\frac{2k}{\mu} + \sqrt{\left(\frac{2k}{\mu}\right)^2 + 1}\right] \rho(T_h^k(\omega_{opt})), \quad k = 1, 2, \cdots,$$

式中 μ 也是 Jacobi 迭代矩阵的谱半径。此式右端第一个因子基本上随 k 线性增长，第二个因子随 k 指数下降。当 k 较小时，第

一因子占主导地位，近似解误差变大；当 k 值较大时，第二因子占主导地位，随着 k 值的增加，近似解误差单调下降，且获得最快的收敛速度(相对于其它 ω 而言)。

下面就模型问题的计算实例，将多种迭代近似解的收敛过程表示在图 2.3 中。在通常情况下，ω-Jacobi 方法和 GS 方法的

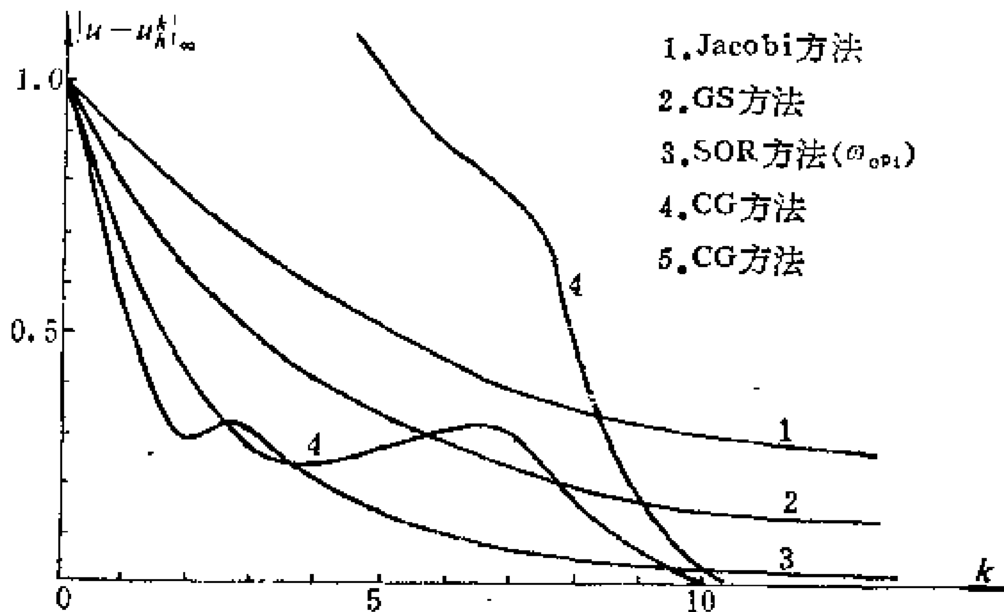


图2.3

近似解 u_h^k 是单调收敛于 u_h 的，初始几步内收敛较快，随着迭代步数的增加，收敛速度变得越来越慢。由 SOR 方法得到的近似值 u_h^k 在初始迭代步内有摆动现象，但后来也是单调收敛于 u_h 的，虽然它的敛速比 Jacobi 方法和 GS 方法快，但随着迭代步数的增加，敛速也是越来越慢。上述三种方法都密切依赖于迭代初始值，初始值好能使计算量大大减少。CG 方法在初始几个迭代步内，近似解收敛得并不快，但较后迭代步中其敛速大增，CG 方法对初始值的敏感性不强。

因此，人们不难想到，有效地利用某些迭代法的初期好的收敛性，有效地利用另一些迭代法后期好的收敛性，将几种迭代方法组合使用，可期望产生较好的收敛效果。下面介绍几种有效的

组合方案。

第一种组合方案。初始几步迭代采用 GS 方法，以后恒用带有最佳松弛因子的 SOR 方法。这样会避免初始迭代过程中SOR误差的增大现象，从而提高了迭代过程的效率。

第二种组合方案。对红黑分划交替使用 GS 和带有最佳松弛因子的 SOR 方法。例如当 $i+j$ 等于奇数时用 GS 方法，$i+j$等于偶数时用SOR方法。可以证明，这种组合迭代的误差模量是单调下降的。

第三种组合方案。设想把红黑分划每个完整迭代步（有二个部分步）当作两次迭代，在 $i+j=$奇或偶的交替迭代过程中，每次都采用不同的迭代参数ω_1，ω_2，…。例如，对模型问题的 5点离散格式，采用下面计算公式

$$
\begin{cases}
u_{i,j}^{2k+1}=u_{i,j}^{2k-1}+\dfrac{\omega_{2k+1}}{4}(f_{i,j}+u_{i-1,j}^{2k}+u_{i+1,j}^{2k}+u_{i,j-1}^{2k} \\
\qquad\qquad +u_{i,j+1}^{2k}-4u_{i,j}^{2k-1}),\quad i+j=\text{奇}, \\
u_{i,j}^{2k+2}=u_{i,j}^{2k}+\dfrac{\omega_{2k+2}}{2}(f_{i,j}+u_{i-1,j}^{2k+1}+u_{i+1,j}^{2k+1}+u_{i,j-1}^{2k+1} \\
\qquad\qquad +u_{i,j+1}^{2k+1}-4u_{i,j}^{2k}),\quad i+j=\text{偶}, \\
\omega_1=1,\quad \omega_2=\dfrac{2}{2-\mu^2},\quad \omega_{k+1}=\dfrac{4}{4-\mu^2\omega_k},\quad k=1,\ 2,\ \cdots,
\end{cases}
$$

式中 μ 是逐点 Jacobi 迭代矩阵的谱半径。这个算法实际上就是所谓的**切比雪夫加速法**。当 $k\to\infty$时，ω_k 稳定到一个极限值，用 $\omega_{k+1}=\omega_k$ 代入上式，即知

$$
\omega_k\to\frac{2}{1+\sqrt{1-\mu^2}}=\omega_{opt}.
$$

因此，这种算法的渐近收敛速度与优选松弛法相同，此方法的初期收敛性也好，因此该方法在实践中也是可取的。关于切比雪夫迭代及切比雪夫加速法的有关推导，有兴趣的读者可参见文献[3]。

第三章　松弛迭代的两个基本特性

为了发扬基本迭代法的优点，克服其缺点，为了研究和构造新的高效率迭代方法，本章对基本迭代法的基本特性作了统一分析。§1中分析了迭代过程的误差校正特性，§2中分析了松弛迭代的光滑特性。正是这两个基本特性的巧妙应用，才产生了第四章中将要介绍的新的高效率迭代方法。

§1　迭代过程的误差校正特性

记 u_h^k 为迭代法解线性离散方程

$$L_h u_h = f_h \tag{3.1}$$

的任一近似值，**误差** $v_h^k = u_h - u_h^k$ 可看作是对 u_h^k 的**校正量**，显然

$$L_h v_h^k = f_h - L_h u_h^k$$

成立。称 $d_h^k = f_h - L_h u_h^k$ 为**剩余向量**或**残差向量**，那么误差 v_h^k 必满足**残差方程**：

$$L_h v_h^k = d_h^k.$$

从而，求解线性离散方程(3.1)的过程可表示为

$$\begin{cases} u_h = u_h^k + v_h^k, \\ L_h v_h^k = d_h^k. \end{cases}$$

若在解残差方程时用简单的矩阵 $\bar{L}_h$ 代替 L_h，即利用求解形式简单且易于计算的残差方程

$$\bar{L}_h \bar{v}_h^k = d_h^k,$$

可获得 u_h^k 的一个近似校正量 $\bar{v}_h^k$，从而可得到 u_h 的一个新的近似值

$$u_h^{k+1} = u_h^k + \bar{v}_h^k;$$

于是，上述作法定义了一个迭代过程

$$\begin{cases} u_h^{k+1} = u_h^k + \bar{v}_h^k, \\ \bar{L}_h v_h^k = d_h^k, \quad k = 0,\ 1,\ \cdots. \end{cases} \tag{3.2}$$

易知

$$u_h^{k+1}=u_h^k+\bar{L}_h^{-1}d_h^k=(I_h-\bar{L}_h^{-1}L_h)u_h^k+\bar{L}_h^{-1}f_h,$$

$$\begin{aligned}\bar{v}_h^{k+1}&=\bar{L}_h^{-1}d_h^{k+1}=\bar{v}_h^k+\bar{L}_h^{-1}(d_h^{k+1}-\bar{L}_h\bar{v}_h^k)\\&=\bar{v}_h^k+\bar{L}_h^{-1}(f_h-L_hu_h^{k+1}-f_h+L_hu_h^k)\\&=(I_h-\bar{L}_h^{-1}L_h)\bar{v}_h^k,\end{aligned}$$

$$\begin{aligned}d_h^{k+1}&=f_h-L_hu_h^{k+1}=f_h-L_h(u_h^k+\bar{v}_h^k)\\&=d_h^k-L_h\bar{v}_h^k=d_h^k-L_h\bar{L}_h^{-1}d_h^k=(I_h-L_h\bar{L}_h^{-1})d_h^k.\end{aligned}$$

由此可知，由(3.2)式定义的迭代过程给出了

近似解渐近收敛因子 $\rho(I_h-\bar{L}_h^{-1}L_h)$；

误差的渐近收敛因子 $\rho(I_h-\bar{L}_h^{-1}L_h)$；

残差的渐近收敛因子 $\rho(I_h-L_h\bar{L}_h^{-1})$。

由此可见，要使利用误差校正办法确定的迭代过程(3.2)收敛且使用方便，关键是选取简单而又便于计算的算子 $\bar{L}_h:\mathbf{R}^N\to\mathbf{R}^N$，且满足 $\rho(I_h-\bar{L}_h^{-1}L_h)<1$，$\bar{L}_h$的好坏将决定迭代过程(3.2)的优劣。

值得注意的是，通常的迭代法都具有这种误差校正的特性，其迭代过程都可用(3.2)式来描述。

例如，ω-Jacobi 迭代，由(2.12)式知

$$u_h^{k+1}=u_h^k+\omega(L_h^d)^{-1}(f_h-L_hu_h^k),$$

于是有

$$\begin{cases}u_h^{k+1}=u_h^k+\bar{v}_h^k,\\L_h^d\bar{v}_h^k=\omega(f_h-L_hu_h^k).\end{cases}$$

所以 ω-Jacobi 迭代方法使用了简单算子 L_h^d 代替复杂算子 L_h，从而求得 u_h^k的近似校正量 $\bar{v}_h^k$。在 ω-Jacobi 方法中

近似解 u_h^k的渐近收敛因子 $\rho(I_h-\omega(L_h^d)^{-1}L_h)$，

校正量 $\bar{v}_h^k$的渐近收敛因子 $\rho(I_h-\omega(L_h^d)^{-1}L_h)$，

残差量 d_h^k的渐近收敛因子 $\rho(I_h-\omega L_h(L_h^d)^{-1})$。

又如 GS 方法，由(2.15)式知

$$u_h^{k+1}=u_h^k+(L_h^d+L_h^l)^{-1}(f_h-L_hu_h^k),$$

于是有

$$\begin{cases}u_h^{k+1}=u_h^k+\bar{v}_h^k,\\(L_h^d+L_h^l)\bar{v}_h^k=f_h-L_hu_h^k.\end{cases}$$

所以在 GS 方法中，使用了简单算子$(L_h^d+L_h^l)$定义残差方程，从而较方便地获得了 u_h^k的近似校正量 $\bar{v}_h^k$。在 GS 方法中

近似解 u_h^k的渐近收敛因子　$\rho(I_h-(L_h^d+L_h^l)^{-1}L_h)$，

校正量 $\bar{v}_h^k$的渐近收敛因子　$\rho(I_h-(L_h^d+L_h^l)^{-1}L_h)$，

残差量d_h^k的渐近收敛因子　$\rho(I_h-L_h(L_h^d+L_h^l)^{-1})$。

再例如SOR方法，由SOR定义有

$$\begin{aligned}u_h^{k+1}&=(1-\omega)u_h^k+\omega(I_h-(L_h^d+L_h^l)^{-1})u_h^k+\omega(L_h^d+L_h^l)^{-1}f_h\\&=u_h^k+\omega(L_h^d+L_h^l)^{-1}(f_h-L_hu_h^k),\end{aligned}$$

于是SOR方法又可改写为

$$\begin{cases}u_h^{k+1}=u_h^k+\bar{v}_h^k,\\(L_h^d+L_h^l)\bar{v}_h^k=\omega(f_h-L_hu_h^k),\end{cases}$$

所以SOR迭代过程所确定的

近似解u_h^k的渐近收敛因子　$\rho(I_h-\omega(L_h^d+L_h^l)^{-1}L_h)$，

校正量$\bar{v}_h^k$的渐近收敛因子　$\rho(I_h-\omega(L_h^d+L_h^l)^{-1}L_h)$，

残差量d_h^k的渐近收敛因子　$\rho(I_h-\omega L_h(L_h^d+L_h^l)^{-1})$。

在下章介绍的多层网格方法中，迭代过程的误差校正特性仍被有效地利用，而且残差方程是确定在低维的网格函数空间中的，残差方程的系数矩阵变得更简单，从而求近似解校正量的工作量更少。

§2　松弛迭代的光滑特性

2.1　ω-Jacobi 迭代的光滑特性分析

松弛迭代过程除了有误差校正特性外，还有**光滑特性**。所谓光滑作用，是指松弛迭代过程对高频误差分量的强有力的衰减作用。为了说明问题，设矩形区域上定义的 Poisson 问题的5点格

式离散方程为

$$L_h u_h = f_h, \quad x \in \Omega_h,$$

其中，$\Omega_h = \{(ih, jh) | i、j = 0, 1, \cdots, N\}$；并以$\omega$-Jacobi松驰迭代过程为例。

先在有限维网格函数空间分析ω-Jacobi松弛迭代过程对各种频率的特征向量的衰减作用。因为5点格式离散方程ω-Jacobi迭代矩阵的特征值为

$$\lambda_n(\omega) = 1 - \frac{\omega}{2}(2 - \cos n_1 \pi h - \cos n_2 \pi h), \tag{3.3}$$

其相应的特征向量为

$$\varphi_n(x) = 2\sin(n_1 \pi x_1)\sin(n_2 \pi x_2), \quad x \in \Omega_h, \tag{3.4}$$

其中，$x = (x_1, x_2)$，$n = (n_1, n_2)$，$|n| = \max\{|n_1|, |n_2|\}$，$1 \leqslant |n| \leqslant N - 1$，$N$是偶数。

由第二章§3知道，ω-Jacobi迭代的敛速为$O(h^2)$。敛速与h的相关性表明：网格步长大，迭代收敛速度较快，但数值解精度低；网格细密，h较小，虽然数值解精度可得到改善，但因迭代算子的谱半径$(1-O(h^2))$更接近于1，所以要达到足够精度不仅会付出很大的工作量，而且当h很小时计算过程会出现数值不稳定。另一方面，迭代前后的近似解误差v_h和$\bar{v}_h$可用特征函数展开为

$$v_h = \sum_{|n| \leqslant N-1} c_n \varphi_n, \qquad \bar{v}_h = \sum_{|n| \leqslant N-1} \lambda_n c_n \varphi_n,$$

其中λ_n和φ_n由(3.3)式和(3.4)式定义，c_n是系数；迭代过程对误差压缩效果及迭代敛速通常是用谱半径来衡量的，即用特征值中绝对值最大的$\lambda_{1,1}$来衡量的，这个特征值对应着最低频率的特征向量$\varphi_{1,1}$，所以敛速快慢通常是指最低频率的特征向量在迭代过程中的衰减程度，它是与h有关的；迭代过程对误差中各种不同频率分量是否存在着某些特殊的衰减作用呢？

为此，定义全部特征向量φ_n由**高频分量**和**低频分量**两部分组成，即

$$\varphi_n = \varphi_n^{\text{low}} \cup \varphi_n^{\text{high}}, \tag{3.5a}$$

其中，

$$\begin{cases} \varphi_n^{\text{low}} = \{\varphi_n \mid |n| < N/2\}, \\ \varphi_n^{\text{high}} = \{\varphi_n \mid N/2 \leqslant |n| \leqslant N-1\}. \end{cases} \tag{3.5b}$$

例如，一维区域$[a, b]$均分成 4 份时，其迭代算子有 3 个不同频率的特征函数 $\varphi_n = \sin(n\pi x)$，$(n=1, 2, 3)$；当均分为 8 份时，迭代算子有 7 个不同频率的特征函数 $\varphi_n = \sin(n\pi x)$，$n=1, 2, \cdots, 7$，而且它的低频分量部分与$[a, b]$ 均分为 4 份时的特征函数一样，它的高频部分则在$[a, b]$均分为 4 份时是没有表现的。

我们还定义**光滑因子**是，经一次迭代步使高频误差分量衰减得最慢的因子，光滑因子及其上确界分别记为$\mu(h, \omega)$和$\mu^*(\omega)$，显然

$$\begin{cases} \mu(h, \omega) = \max\{|\lambda_n| \mid N/2 \leqslant |n| \leqslant N-1\}, \\ \mu^*(\omega) = \sup\{\mu(h, \omega) \mid h \leqslant 1/4\}. \end{cases} \tag{3.6}$$

由(3.3)式～(3.6)式可知，ω-Jacobi 松弛法解 Poisson 问题时的光滑因子及其上确界为

$$\begin{cases} \mu(h, \omega) = \max\left\{\left|1 - \dfrac{\omega}{2}(2 - \cos\pi h)\right|, \ |1 - \omega(1 + \cos\pi h)|\right\}, \\ \mu^*(\omega) = \max\left\{\left|1 - \dfrac{\omega}{2}\right|, \ |1 - 2\omega|\right\}. \end{cases}$$

当 $0 < \omega \leqslant 1$ 时，有

$$\mu(h, \omega) = \begin{cases} \cos\pi h, \text{ 当 } \omega = 1 \text{ 时}, \\ \dfrac{(2 + \cos\pi h)}{4}, \text{ 当 } \omega = \dfrac{1}{2} \text{时}, \\ \dfrac{(1 + 2\cos\pi h)}{5}, \text{ 当} \omega = \dfrac{4}{5} \text{时}, \end{cases} \tag{3.7}$$

$$\mu^*(\omega) = \begin{cases} 1, \text{ 当 } \omega = 1 \text{ 时}, \\ \dfrac{3}{4}, \text{ 当 } \omega = \dfrac{1}{2} \text{时}, \\ \dfrac{3}{5}, \text{ 当 } \omega = \dfrac{4}{5} \text{时}. \end{cases} \tag{3.8}$$

特别地，

$$\inf\{\mu(h,\ \omega)\,|\,0<\omega\leqslant 1\}=\mu\left(h,\ \frac{4}{4+\cos\pi h}\right)$$

$$=\frac{3\cos\pi h}{4+\cos\pi h}=\frac{3}{5}-O(h^2),$$

$$\inf\{\mu^*(\omega)\,|\,0<\omega\leqslant 1\}=\mu^*\left(\frac{4}{5}\right)=\frac{3}{5}.$$

局部Fourier 分析法能更简捷地描述和分析松弛迭代过程的光滑特性。为此，只要把有限网格 Ω_h 扩展成无限网格G_h，按第二章§2中的办法求得离散算子 L_h 的特征函数 $\varphi(h,\ x)=\exp(i\pi lx)$，再令 $\pi lh=\theta$，则可将特征函数改写为

$$\varphi(\theta,\ x)=\exp(i\theta x/h),\quad \theta\in[-\pi,\ +\pi],\quad x\in G_h.$$

$\varphi(\theta,\ x)$也是在 G_h 上确定的网格函数空间的基函数，它是由若干频率函数组成的，其中有低频函数，也有高频函数。为了用局部 Fourier 方法研究迭代过程对不同频率函数的衰减作用，仿(3.5a)和(3.5b)式定义 $\exp(i\theta x/h)$中的所有频率函数分成低频量和高频量两部分，即

$$\varphi(\theta,\ x)^{\mathrm{low}}=\left\{\varphi(\theta,\ x)\,\middle|\,|\theta|<\frac{\pi}{2}\right\},$$

$$\varphi(\theta,\ x)^{\mathrm{high}}=\left\{\varphi(\theta,\ x)\,\middle|\,\frac{\pi}{2}\leqslant|\theta|\leqslant\pi\right\}.$$

不难理解，若在步长为 $2h$ 的网格 G_{2h} 基础上细分一次得到步长为 h 的网格 G_h，则在 G_{2h} 和 G_h 上定义的离散算子 L_{2h} 和 L_h 的特征函数可分别表示为

$$\Phi(\theta,\ x)=\exp(i\theta x/2h),\quad \theta\in\left(-\frac{\pi}{2},\ \frac{\pi}{2}\right),\quad x\in G_{2h},$$

$$\varphi(\theta,\ x)=\exp(i\theta x/h),\quad \theta\in[-\pi,\ \pi],\quad x\in G_h,$$

其中$\Phi(\theta,\ x)$中 θ 的变化范围$\left(-\frac{\pi}{2},\ \frac{\pi}{2}\right)$是相对于$\varphi(\theta,\ x)$中的 $\theta\in[-\pi,\ \pi]$而言的，显然

$$\varphi(\theta,\ x)=\Phi(\theta,\ x),\quad \forall x\in G_{2h}.$$

所以，$\varphi(\theta, x)$中的低频分量部分与G_{2h}上的$\Phi(\theta, x)$的全体频率分量是等同的，但$\varphi(\theta, x)$中的高频分量部分是在G_{2h}上没有表现的。

这样，光滑因子的定义(3.6)式又可改写为

$$\mu(\theta, \omega) = \max\left\{ |\lambda(\theta, \omega)| \,\middle|\, \frac{\pi}{2} \leqslant \theta \leqslant \pi \right\},$$

$$\mu^*(\omega) = \sup\{\mu(\theta, \omega) | \theta \to 0\}.$$

因此，只要利用定理2.2求出5点离散格式ω-Jacobi迭代算子的特征值

$$A(\theta, \omega) = 1 - \frac{\omega}{2}(2 - \cos\theta_1 - \cos\theta_2),$$

再利用光滑因子的定义，就可得到

$$\mu(\theta, \omega) = \max\left\{ |A(\theta, \omega)| \,\middle|\, |\theta| \in \left[\frac{\pi}{2}, \pi\right] \right\}$$

$$= \max\left\{ \left|1 - \frac{\omega}{2}(2 - \cos\theta)\right|, |1 - \omega(1 + \cos\theta)| \right\},$$

$$\mu^*(\omega) = \sup_{\theta \to 0} \mu(\theta, \omega) = \max\left\{ \left|1 - \frac{\omega}{2}\right|, |1 - 2\omega| \right\}$$

$$= \begin{cases} 1, & \omega = 1, \\ 3/4, & \omega = 1/2, \\ 3/5, & \omega = 4/5. \end{cases}$$

由此可见，用局部Fourier分析方法分析ω-Jacobi迭代的光滑特性的结果是与在有限维空间中分析光滑特性的结果一致的。

现将ω-Jacobi迭代矩阵的特征值随着频率nh和松驰因子ω变化的情况表示于图3.1中；将经过ω-Jacobi迭代的$u_h^k - u_h (k = 0, 1, 2,)$示意于图3.2中，还将$\omega$-Jacobi迭代对低频率分量压缩和对高频率分量光滑的效果示意于图3.3中。

本节的数值分析结果和图3.1～图3.3表明：当$0 < \omega \leqslant 1$时，ω-Jacobi方法对误差的压缩作用是明显的，参见图3.1和图3.2，经迭代作用后近似解越来越逼近准确解；ω-Jacobi迭代对误差

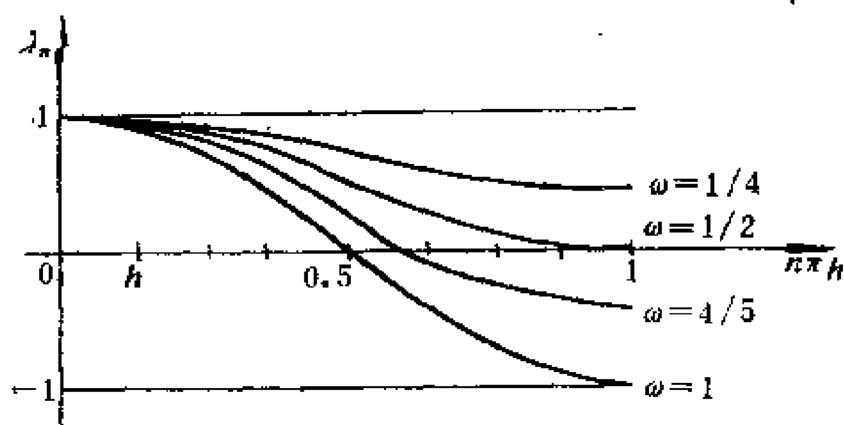


图 3.1

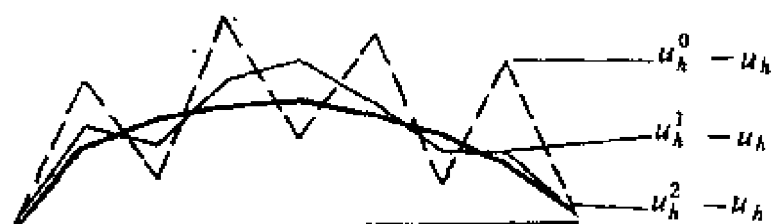


图 3.2

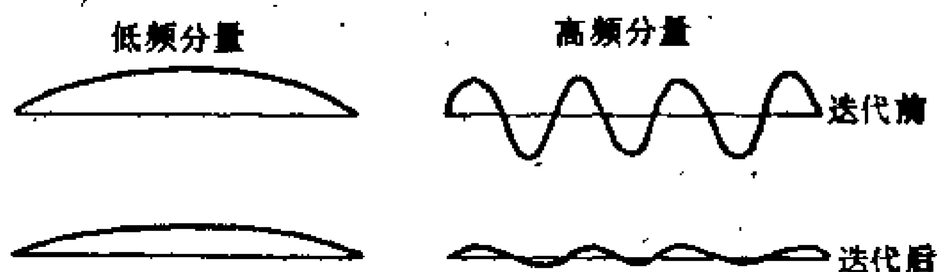


图 3.3

的低频分量衰减缓慢且当 $h\to 0$ 时其衰减效果越差，参见图3.1和图3.3，据谱半径定义，迭代的整体收敛因子实际上就是迭代对低频分量的收敛因子，所以低频量收敛慢是整体迭代收敛缓慢的本质原因；ω-Jacobi 迭代对误差的高频分量具有显著的压缩效果，参见图3.1、图3.3和(3.7)式及(3.8)式；当 $\omega=1$ 时，$\mu(1)=1-O(h^2)$；当 $\omega=4/5$ 时，$\mu(4/5)=3/5-O(h^2)$；特别有意思且令人兴奋的是当 $0<\omega<1$ 时，ω-Jacobi 方法对高频分量的光滑

效果可以 h 无关地小于 1（参见(3.8)式），且当 $\omega=4/5$ 时光滑效果最佳，即使网格步长 h 充分小也有 $\mu^*(4/5)=3/5$。对高频误差分量具有高效率且 h 无关的光滑作用正是松弛迭代过程的本质特点之一，这一特点将被"多层网格方法"所引用并产生了 h 无关收敛性的迭代效果。

2.2 逐点GS迭代的光滑特性分析

考虑二阶椭圆边值问题

$$\begin{cases} Lu = a\partial^2 u/\partial x^2 + b\partial^2 u/\partial y^2 = f, & (x,\ y)\in\Omega, \\ u|_{\partial\Omega}=0, \end{cases} \tag{3.9}$$

若采用正方形网格分划，步长为 h；节点坐标式编号；5 点离散格式；则有离散方程

$$L_h u_{ij} = \frac{a}{h^2}(u_{i-1,j} - 2u_{ij} + u_{i+1,j}) + \frac{b}{h^2}(u_{i,j-1} - 2u_{ij} + u_{i,j+1}) = f_{ij}. \tag{3.10}$$

(3.10)式的逐点GS迭代公式为

$$a(u_{i-1,j}^{k+1} - 2u_{ij}^{k+1} + u_{i+1,j}^{k}) + b(u_{i,j-1}^{k+1} - 2u_{ij}^{k+1} + u_{i,j+1}^{k}) = h^2 f_{ij}, \tag{3.11}$$

此时，离散格式分解及相应的特征值表示为

$$[L_h^+] = \begin{pmatrix} 0 & 0 & 0 \\ a & 0 & 0 \\ 0 & b & 0 \end{pmatrix}, \qquad [L_h^0] = \begin{pmatrix} 0 & 0 & 0 \\ 0 & -2(a+b) & 0 \\ 0 & 0 & 0 \end{pmatrix},$$

$$[L_h^-] = \begin{pmatrix} 0 & b & 0 \\ 0 & 0 & a \\ 0 & 0 & 0 \end{pmatrix}, \qquad \begin{aligned} \lambda^+ &= a\exp(-i\theta_1) + b\exp(-i\theta_2), \\ \lambda^0 &= -2(a+b), \\ \lambda^- &= a\exp(i\theta_1) + b\exp(i\theta_2), \end{aligned}$$

从而得GS局部迭代算子的特征值 $|A(\theta)| = |\lambda^-/(\lambda^0+\lambda^+)|$。

当 $a=b$ 时，(3.9)式是 Poisson 问题，GS 方法的光滑因子为

$$\mu = \max\{|A(\theta_1,\ \theta_2)|\},\quad |\theta|\in\left[\frac{\pi}{2},\ \pi\right],$$

$$\mu^* = \mu\left(\frac{\pi}{2}, \quad \arccos\frac{4}{5}\right) = 0.5.$$

此时，GS 方法具有良好的光滑作用。

当 $a \ll b$（或 $b \ll a$）时，有

$$\mu^2\left(\frac{\pi}{2}, 0\right) = [a^2 + b^2]/[a^2 + (2a+b)^2], \ \mu^* \to 1 \quad (a \to 0).$$

此时，GS方法的光滑作用很差。

2.3 逐线GS方法的光滑特性分析

对于问题(3.10)，采用 y 方向的逐线GS 方法，其计算公式为：

$$\frac{a}{h^2}(u_{i-1,j}^{k+1} - 2u_{ij}^{k+1} + u_{i+1,j}^{k}) + \frac{b}{h^2}(u_{i,j-1}^{k+1} - 2u_{ij}^{k+1} + u_{i,j+1}^{k+1}) = f_{ij}. \tag{3.12}$$

此时，$\lambda^+ = a\exp(-i\theta_1)$，$\lambda^0 = 2b\cos\theta_2 - 2(a+b)$，$\lambda^- = a\exp(i\theta_1)$，于是利用定理2.2求得 $A(\theta) = -\lambda^-/(\lambda^0 + \lambda^+)$，从而求出光滑因子的上界

$$\mu^* = \max|A(\theta)|^{1/2} \quad \left(|\theta| \in \left[\frac{\pi}{2}, \pi\right]\right)$$
$$= \max\{1/\sqrt{5}, \ a/[a+2b]\}.$$

由此可知，当$a \ll b$时，y 方向的逐线 GS 迭代具有很好的光滑效果；当$b \ll a$时，y 方向的逐线GS迭代的光滑效果很差，但易知，此时 x 方向线GS迭代的光滑效果是很好的。

总之，当 a、b 都是常数 $a=b$或$a \approx b$时，问题(3.9)为各向同性或弱各向异性的椭圆型问题，此时逐点 GS 方法具有良好的光滑效果；当 a、b 都是常数，$a \ll b$或$b \ll a$ 时，问题(3.9)为强各向异性问题，此时沿 a 或 b 数值较大的方向作逐线 GS 迭代会有良好的光滑效果；当 a、b 都为 x 的函数时，可采用“冻结系数法”，即将$a(x_1, x_2)$和$b(x_1, x_2)$代之以$a(ih, jh)$和$b(ih, jh)$，再利用局部Fourier分析方法研究光滑效果，只不过这样得到的光滑因子只能反应(ih, jh)附近的局部情形而已。

第四章　多层网格方法

构造高效率求解离散方程的迭代法一直是计算数学工作者和工程技术人员共同关心的问题，第二章中所介绍的各种基本迭代法以及为改进这些迭代法所作出的种种努力都是在固定的网格分划下进行的，这些迭代法解离散微分方程的敛速都与网格参数 h 有关，其计算工作量都为$O(N^{\alpha})$，$\alpha\geqslant 1$，N是离散方程中未知量的个数。

60年代初，R. P. Fedorenko 提出了求解离散方程的一种新的迭代法[38]，称为多层网格方法。70年代初，A. Brandt 进一步总结研究和发展了多层网格方法。后来，许多计算数学家对多层网格方法的理论和应用问题作了深入的研究。多层网格方法有效地利用了迭代过程的误差校正特性和对高频误差分量的光滑特性，改变了传统的作法，作了变革性的构造。多层网格方法的敛速与 h 无关，其计算工作量为 $O(N)$，多层网格方法是一种高效率迭代法。

本章不仅介绍了多层网格方法的基本思想、基本结构，给出了其实现过程解析和计算框图，而且分析了多层网格方法的 h 无关收敛性和工作量估计，本章中还给出了多种数值方法与多层网格方法的算例数值比较。

§1　多层网格方法的基本思想

按通常的习惯，数值求解微分方程的过程是这样的，先将定解区域Ω用某种方法离散为Ω_h，h 是其网格参数；再建立定义在Ω_h上的离散微分方程，并求其解，然而，求解过程却不再与Ω_h发生什么联系。虽然在细的网格层Ω_h上求解离散方程会得到较高精

度的数值解，但工作量花费大；虽然在粗的网格分划层 $\Omega_H(H\gg h)$ 上所得到的数值解精度较差，但求解工作量要比细网格层上的工作量小得多。因此，人们自然地想到，能否将细网格层上的数值求解问题转化为粗网格层上的数值求解问题，从而节省计算量，提高工作效率呢？

1.1 粗网格校正

第三章§1叙述了迭代方法的误差校正特性，并以此定义了迭代过程的一般准则，若校正量由粗网格层上的信息提供，即(3.2)式中校正量满足的残差方程定义在粗网格层上，则会使我们得到启发。

为此，设求解的离散微分方程形如

$$L_h u_h = f_h,\quad u_h\in G(\Omega_h),\tag{4.1}$$

其中，$G(\Omega_h)$表示定义在Ω_h上的网格函数空间；若无特殊说明和需要，不妨设(4.1)式就是差分方程；再设

$$L_H u_H = f_H,\quad u_H\in G(\Omega_H)\tag{4.2}$$

与(4.1)式相应，是同一微分方程的离散方程，只不过分划 Ω_H比Ω_h粗；显然

$$L_h:\ G(\Omega_h)\to G(\Omega_h),\quad L_H:\ G(\Omega_H)\to G(\Omega_H),$$
$$\dim G(\Omega_H)\ll \dim G(\Omega_h),\quad H\ll h,$$

其中 $\dim G(\Omega_h)$ 表示 Ω_h中的节点个数或方程组(4.1)中的未知量的个数；并设 d_h、v_h 与 d_H、v_H 分别是求解(4.1)式和(4.2)式时出现的残差与误差；还设两个联系粗细网格函数的算子

$$I_h^H:\ G(\Omega_h)\to G(\Omega_H),\quad I_H^h:\ G(\Omega_H)\to G(\Omega_h),$$

其中，I_h^H称为**限制算子**，I_H^h称为**内插算子**。

为了数值求解 (4.1)式，利用误差校正特性定义的迭代过程可描述为

$$u_h = u_h^k + v_h^k,$$
$$L_h v_h^k = d_h^k,$$

其中，u_h^k是任给的近似值，(4.2)是细网格层上的残差方程。

为了能在粗网格层上计算，需将(4.2)式变换到粗网格层上，利用限制算子I_h^H就有

$$\tilde{L}_H v_H^k = d_H^k,$$

式中，$\tilde{L}_H = I_h^H L_h$，$v_H^k = I_h^H v_h^k$，$d_H^k = I_h^H d_h^k$。然而 $\tilde{L}_H$ 可能是复杂的，所以可用L_H代替$\tilde{L}_H$，即可得到粗网格层上的残差方程

$$L_H v_H^k = d_H^k. \tag{4.3}$$

解出粗网格层上的校正量

$$v_H^k = L_H^{-1} d_H^k = L_H^{-1} I_h^H d_h^k, \tag{4.4}$$

再将v_H^k内插到细网格层，得

$$\hat{v}_h^k = I_H^h v_H^k. \tag{4.5}$$

最后利用$\hat{v}_h^k$作为细网格层上近似值u_h^k的校正量，从而得到一个新的近似值

$$u_h^{k+1} = u_h^k + \hat{v}_h^k. \tag{4.6}$$

总而言之，利用求解粗网格层(Ω_H)上的残差方程(4.3)，可以求得细网格层(Ω_h)上差分方程新的近似值，这一过程称为**粗网格校正过程**。因为求解粗网格层上残差方程的工作量较小，所以，如果粗网格校正过程收敛，就必将加快求解(4.1)式的速度。粗网格校正过程是多层网格方法的基本思想之一。

粗网格校正过程的具体步骤如下：

(1) 计算残差量，$d_h^k = f_h - L_h u_h^k$；

(2) 限制残差量，$d_H^k = I_h^H d_h^k$；

(3) 在Ω_H上解残差方程，$L_H v_H^k = d_H^k$；

(4) 内插校正量，$\hat{v}_h^k = I_H^h v_H^k$；

(5) 计算新的近似值，$u_h^{k+1} = u_h^k + \hat{v}_h^k$。

粗网格校正过程可简单地图示如下：

$$\begin{array}{lccc}
\Omega_h & u_h^k \to d_h^k = f_h - L_h u_h^k & & \hat{v}_h^k \to u_h^{k+1} = u_h^k + \hat{v}_h^k \\
 & \downarrow I_h^H & & \uparrow I_H^h \\
\Omega_H & d_H^k & \longrightarrow & L_H v_H^k = d_H^k
\end{array}$$

粗网格校正过程是一个迭代过程，其迭代算子是

$$K_h^H = I_h - I_H^h L_H^{-1} I_h^H L_h. \tag{4.7}$$

事实上，

$$u_h^{k+1} = u_h^k + \hat{v}_h^k = u_h^k + I_H^h v_H^k = u_h^k + I_H^h L_H^{-1} d_H^k$$
$$= u_h^k + I_H^h L_H^{-1} I_h^H (f_h - L_h u_h^k).$$

1.2 误差光滑

仅利用粗网格校正过程求解离散方程(4.1)的效果可能是很差的，一般得不到足够精度的解，甚至是不收敛的。事实上，如选择初始值 w_h 使得 $L_h w_h$ 属于 I_h^H 的核，那么经粗网格校正算子(4.7)作用后仍然是 w_h 本身。因此，粗网格层上的残差方程(4.3)并不一定能有效地反映细网格层上的残差方程(4.2)。这是因为将 v_h^k 和 d_h^k 限制到粗网格层后，损失了一部分细网格层上的信息，具体地说，仅有上述量的低频分量在Ω_H上是有表现的，而其高频分量在 Ω_H上是没有表现的。如果由Ω_h上的残差方程(4.2)所确定的误差v_h^k是“光滑”的，即如果v_h^k的高频分量比它的低频分量小得多，则Ω_H上的残差方程(4.3)的解就可以更好地表现 Ω_h 上的残差方程(4.2)的解了。另一方面，由于内插算子I_H^h的作用，v_h^k 内插到细网格层后，又增加了校正量的高频分量，保真度降低，所以粗网格校正后也需要“光滑”.综上所述，在粗网格校正前后，需要对误差的高频分量压缩即“光滑”，才能使粗网格校正方法发挥合乎情理的有效作用。这就需要用到松弛方法对误差高频分量的光滑作用了，这是多层网格方法的另一个基本思想。

§2 两层网格方法

2.1 两层网格方法的结构

综合粗网格校正和松弛光滑这两个过程，可以确定一个切实可行的定义在Ω_h和Ω_H上的两层网格方法.

两层网格方法解(4.1)式的一个完整迭代步骤具体如下：

(1) 粗网格校正前光滑。任给 Ω_h 网格层上的初值 u_h^k，并对(4.1)式施行 ν_1 步松弛迭代，得新的近似值 $\bar{u}_h^k$，即

$$\bar{u}_h^k = \mathrm{Relax}(u_h^k, L_h, f_h, \nu_1);$$

(2) 粗网格校正过程。

计算并限制残差 $d_H^k = I_h^H d_h^k = I_h^H(f_h - L_h\bar{u}_h^k)$,

解出残差方程 $L_H v_H^k = d_H^k$,

内插校正量 $v_h^k = I_H^h v_H^k$,

计算校正后的近似值 $\hat{u}_h^k = \bar{u}_h^k + v_h^k$;

(3) 粗网格校正后光滑。以 $\hat{u}_h^k$ 为新近似值，在 Ω_h 网格层上对(4.1)式施行 ν_2 步迭代，得 u_h^{k+1}，即

$$u_h^{k+1} = \mathrm{Relax}(\hat{u}_h^k, L_h, f_h, \nu_2).$$

两层网格方法一个迭代步的结构可简单图示如下：

$$\begin{array}{ccc} u_h^k \xrightarrow[\mathrm{Relax}]{\nu_1} \bar{u}_h^k \to d_h^k = f_h - L_h\bar{u}_h^k & & v_h^k \to \bar{u}_h^k + v_h^k \xrightarrow[\mathrm{Relax}]{\nu_2} u_h^{k+1} \\ \downarrow I_h^H & & \uparrow I_H^h \\ d_H^k \text{---} & \longrightarrow & L_H v_H^k = d_H^k \end{array}$$

由前所述，立即可得到两层网格方法的迭代算子

$$\begin{aligned} M_h^H &= S_h^{\nu_2} K_h^H S_h^{\nu_1} \\ &= S_h^{\nu_2}(I_h - I_H^h L_H^{-1} I_h^H L_h) S_h^{\nu_1}, \end{aligned} \tag{4.8}$$

其中 K_h^H 是同于(4.7)式的粗网格校正算子，S_h 为粗网格校正前后所采用同一松弛迭代算子。如果 M_h^H 的谱半径小于1，则两层网格方法收敛，收敛性证明参见[12～13]和本章§4.

长期来，加快迭代方法的收敛速度，减少计算量，研究和改进更好的迭代方法一直是倍受重视的。例如，第二章中的GS、LGS和RB方法等就是通过改变 $[L_h] = [L_h^+] + [L_h^0] + [L_h^-]$ 的分解形式达到了改进迭代收敛速度的目的；例如SOR方法就是利用引进加速因子达到改进迭代敛速的目的；又如[3]中介绍了一种多项式加速法，它将基本迭代法得到的迭代值进行线性组合，

形成一个新的迭代序列。所有这些加速方法都局限在Ω_h网格层上进行，计算工作量仍是很大，且敛速总是与 h 有关。两层网格方法以及后面将介绍的多层网格方法改变了传统的作法：它将细网层上求解校正量的问题转化为在粗网层上(甚至更粗的网格层上)进行，为减少计算量作了变革性的构造；它还利用松弛方法的光滑特性，使该方法敛速与 h 无关，为迭代的收敛性作了革命性的改造。因此，该方法发扬了基本迭代法的优点，克服了基本迭代法的缺点，是一种很有生命力的数值方法。

2.2 两层网格方法中的参数选择

由两层网格方法的定义可知，两层网格方法是与多个参数有关的。如网格分划方式，粗网层上残差方程算子的选择、内插和限制算子的定义等，这些将在下面介绍。

1. 粗网格Ω_H的选取

这里仅就矩形网格剖分情形，列举三种常用的网格剖分。记$\Omega_h=\Omega\cap\{(ih_1,jh_2)|i,j=1,2,\cdots\}$，$h=(h_1,h_2)$。

若$H=2h=(2h_1,2h_2)$，则称Ω_H为**标准粗化网格**，标准粗化网格是常用的。参见图4.1a)。

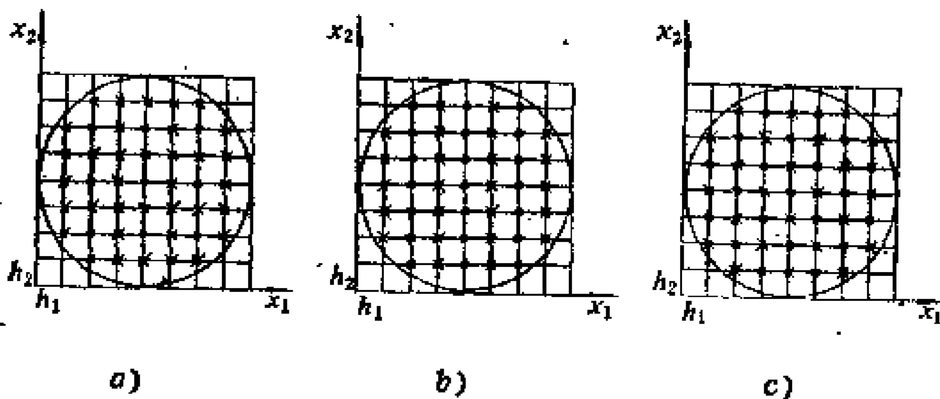


图4.1 Ω_h由"*"和"·"点组成，Ω_H由"·"点组成

若$H=(2h_1,h_2)$或$H=(h_1,2h_2)$，则称Ω_H为**半粗化网格**。参见图4.1b)。它对采用逐线松弛光滑是适用的，逐线松弛法对各项异性微分算子相应的差分方程具有很好的光滑作用，参见第

三章§2.3.

红黑粗化网格Ω_H也是常用的，此时Ω_H的网格点以棋盘分划方式分布在细网格Ω_h的节点上，参见图4.1(c)。显见，当Ω_h是正方形网格时，将其旋转$\pi/4$并把网格步长放大到$\sqrt{2}h$就得Ω_H。

2. 粗网格算子的选择

通常为了方便起见，粗网格算子L_H的构造方式可与细网格算子L_h的构造方式相同。例如L_h采用5点离散格式，L_H也可以采用这种格式。但是原则上，L_H可以与L_h不同，L_H可以是合理逼近L_h的任何一种离散格式。

3. 限制算子的选择

为简单起见，先考虑Ω_H是标准粗化网格，即$H=2h$。限制算子I_h^{2h}的作用是将Ω_h上的网格函数变换为Ω_{2h}上的网格函数，它可以仿照局部离散格式写成如下形式：

$$I_h^{2h}=\begin{pmatrix} r_{-1,1} & r_{0,1} & r_{1,1} \\ r_{-1,0} & r_{0,0} & r_{1,0} \\ r_{-1,-1} & r_{0,-1} & r_{1,-1} \end{pmatrix}_h^{2h}.$$

最常用的限制算子有**全权限制算子**和**直接限制算子**

$$\frac{1}{16}\begin{pmatrix} 1 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 1 \end{pmatrix}_h^{2h} \text{和} \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}_h^{2h}.$$

对于全权限制算子，有

$$\begin{cases} I_h^{2h}u_{ij}^h=u_{ij}^{2h}, \quad (i,j)\in\Omega_h\cap\Omega_{2h}, \\ I_h^{2h}u_{ij}^h=\frac{1}{16}(u_{i-1,j+1}^h+2u_{i,j+1}^h+u_{i+1,j+1}^h+2u_{i-1,j}^h \\ \quad +4u_{ij}^h+2u_{i+1,j}^h+u_{i-1,j-1}^h+2u_{i,j-1}^h+u_{i+1,j-1}^h), \end{cases}$$

其中，u_{ij}^h和u_{ij}^{2h}分别表示u_h和u_{2h}在节点(i,j)处的离散值；对于直接限制算子，有

$$I_h^{2h}u_{ij}^h=u_{ij}^{2h}=u_{ij}^h, \quad (i,j)\in\Omega_h\cap\Omega_{2h},$$

由于限制算子的局部性，所以只要对粗网格节点循环，再将此点

及其在细网格中的邻点处的 u_h 的函数值按上面公式加权聚集计算，即可实现限制过程。

若采用半粗化网格 Ω_{2h}，参见图4.1(b)，全权限制算子的局部格式形如

$$I_h^{2h}=\frac{1}{4}[1\quad 2\quad 1]_h^{2h},$$

即

$$\begin{cases} I_h^{2h}u_{ij}^h=u_{ij}^{2h}, & (i,j)\in\Omega_h\cap\Omega_{2h}, \\ I_h^{2h}u_{ij}^h=\frac{1}{4}(u_{i-1,j}^h+2u_{ij}^h+u_{i+1,j}^h). \end{cases}$$

若采用红黑粗化网格，此时的限制算子可同于标准粗化情形的限制算子。

值得指出的是，这里介绍的都是线性限制算子，人们当然可以采用高次限制算子，但从计算量角度考虑，线性限制算子是简单易行的。

4. 内插算子的选择

内插算子的作用是将 Ω_H 上的网格函数变换为 Ω_h 上的网格函数，它可以写成局部离散格式。先以标准粗化情形为例，此时内插算子可写成

$$I_{2h}^h=\begin{pmatrix} p_{-1,1} & p_{0,1} & p_{1,1} \\ p_{-1,0} & p_{0,1} & p_{1,0} \\ p_{-1,-1} & p_{0,-1} & p_{1,-1} \end{pmatrix}_{2h}^h.$$

对于 $(i,j)\in\Omega_{2h}$，粗网格函数值 u_{ij}^{2h} 经内插算子作用后，被加权地分配到细网格层上去，即有

$$u_{i+k_1,j+k_2}^h=p_{k_1,k_2}u_{ij}^{2h},\quad k_1、k_2=0,\pm 1.$$

因此，只要对粗网格节点循环，被上面公式计算并注意累加性即可实现内插过程。双线性插值算子

$$I_{2h}^h=\frac{1}{4}\begin{pmatrix} 1 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 1 \end{pmatrix}_{2h}^h$$

是经常使用的。

红黑粗化网格情形的内插算子同于标准粗化情形。对于半粗化网格情形，线性内插算子形如

$$I_{2h}^{h}=\frac{1}{2}[1 \quad 2 \quad 1]_{2h}^{h}.$$

值得指出的是，对2阶微分方程采用线性或双线性内插算子简单易行且计算工作量少，对$2m$阶微分方程应采用m次插值为好。

2.3 两层网格方法的数值分析举例

这里将要进行数值分析的两层网格方法例子中采用下述参数：

(i) $\bar{\Omega}=[0,1]\times[0,1]$被分成$N\times N$等分得$\Omega_h$，$h_1=h_2=1/N$，$N$是偶数，$\Omega_{2h}$为标准粗化网格；

(ii) 考虑Poisson方程，L_h和L_{2h}都采用5点离散格式；

(iii) 采用ω-Jacobi迭代作为光滑方法，光滑算子记为$S_{h;\omega}$，粗网格校正前作ν_1步、校正后作ν_2步光滑；

(iv) 采用全权限制算子和双线性内插算子。

为了在有限维网格函数空间中分析两层网格方法的收敛性，先分析细网格函数空间的结构。

L_h和L_{2h}的特征函数分别为

$$\varphi(n,x)=2\sin(n_1\pi x_1)\sin(n_2\pi x_2),\quad x\in\Omega_h,1\leqslant|n|\leqslant N-1, \tag{4.9}$$

$$\Phi(n,x)=2\sin(n_1\pi x_1)\sin(n_2\pi x_2),x\in\Omega_{2h},\quad 1\leqslant|n|\leqslant\frac{N}{2}-1. \tag{4.10}$$

由第三章§2知，下面两种低频分量的定义是等同的。

$$G(\Omega_h)^{\rm low}=\{\varphi(n,x)\,|\,1\leqslant|n|\leqslant\frac{N}{2}-1,\quad x\in\Omega_h\cap\Omega_{2h}\}$$

$$G(\Omega_{2h})=\{\Phi(n,x)\,|\,1\leqslant|n|\leqslant\frac{N}{2}-1,\quad x\in\Omega_{2h}\},$$

其中

$$G(\Omega_h)=G(\Omega_h)^{\text{low}}\oplus G(\Omega_h)^{\text{high}},$$

$$G(\Omega_h)^{\text{low}}=\{\varphi(n,x)\mid 1\leqslant|n|\leqslant\frac{N}{2}-1,\ x\in\Omega_h\},$$

$$G(\Omega_h)^{\text{high}}=\{\varphi(n,x)\mid \frac{N}{2}\leqslant|n|\leqslant N-1\}.$$

因此，有必要将细网格函数空间$G(\Omega_h)$中的基函数$\varphi(n,x)$进行分类。由于

$$\begin{aligned}\varphi_{N-n_1,n_2}(x)&=2\sin(N-n_1)\pi x_1\sin n_2\pi x_2\\&=2\sin(N\pi x_1-n_1\pi x_1)\sin n_2\pi x_2\\&=2\sin(j\pi-n_1\pi x_1)\sin n_2\pi x_2\\&=\begin{cases}\varphi_{n_1,n_2}(x), & j\text{是奇数},\\-\varphi_{n_1,n_2}(x), & j\text{是偶数},\end{cases}\end{aligned}$$

其中，$x_1=jh_1$（j是正整数），$Nh=1$，$1\leqslant|n|\leqslant\frac{N}{2}$，$N$是偶数；同样，$\varphi_{n_1,N-n_2}(x)$和$\varphi_{N-n_1,N-n_2}(x)$都可能与$\varphi_{n_1,n_2}(x)$不尽相同；所以，当把$\varphi(n,x)$中的低频量$\left(1\leqslant|n|\leqslant\frac{N}{2}-1\right)$连同$|n|=\frac{N}{2}$时的频率量归于一类时，则细网格上的全部特征量至多分为4个部分，并由此张成一个至多是“4维”的细网格函数空间，即

$$E_{h,n}=\text{span}\{\varphi_{n_1,n_2}(x),\ \varphi_{N-n_1,N-n_2}(x),\\ -\varphi_{N-n_1,n_2}(x),\ -\varphi_{n_1,N-n_2}(x)\},\ |n|\leqslant\frac{N}{2}.$$

当$|n|<\frac{N}{2}$时，$E_{h,n}$是“4维”的；当$n_1=\frac{N}{2}$或$n_2=\frac{N}{2}$时，$E_{h,n}$是“2维”的；当$n_1=n_2=\frac{N}{2}$时，$E_{h,n}$是“1维”的。

这样，$E_{h,n}\to E_{h,n}$的变换就对应着一个至多是4×4阶的块对角矩阵；当$|n|<\frac{N}{2}$时，它是4×4阶对角阵；当$n_1=N/2$或$n_2=N/2$时，它是2×2阶对角阵；当$n_1=n_2=N/2$时，它是1×1阶矩阵。这些矩阵中的元素就是该变换与$E_{h,n}$中某个“基底”相对应的特征值。由于迭代算子一般可写成离散算子L_h的多项式形式，迭

代算子就是$E_{h,n}\to E_{h,n}$的变换，所以利用这种变换的特征，可以定量地分析迭代算子以及两层网格算子的特征值。

下面讨论ω-Jacobi 迭代矩阵的特征值。

首先由定理2.2求得5点离散格式 L_h 的特征值

$$\lambda(L_h)=4\left(\sin^2\frac{n_1\pi h}{2}+\sin^2\frac{n_2\pi h}{2}\right),\quad |n|\leqslant N-1,$$

和特征向量 φ_{n_1,n_2}。与$E_{h,n}$的4类基底相对应，用指标$|n|\leqslant\dfrac{N}{2}$ 表示，$\lambda(L_h)$也可分为如下4类：

$$\begin{cases}\varphi_{n_1,n_2}=4\left(\sin^2\dfrac{n_1\pi}{2N}+\sin^2\dfrac{n_2\pi}{2N}\right),\\ \varphi_{N-n_1,N-n_2}=4\left(\cos^2\dfrac{n_1\pi}{2N}+\cos^2\dfrac{n_2\pi}{2N}\right),\\ \varphi_{N-n_1,n_2}=4\left(\cos^2\dfrac{n_1\pi}{2N}+\sin^2\dfrac{n_2\pi}{2N}\right),\\ \varphi_{n_1,N-n_2}=4\left(\sin^2\dfrac{n_1\pi}{2N}+\cos^2\dfrac{n_2\pi}{2N}\right)\end{cases}\tag{4.11}$$

从而5点离散格式ω-Jacobi迭代矩阵适合关系

$$S_{h,n}(\omega)=I_h-\frac{\omega}{4}L_h,$$

$$\begin{aligned}&S_{h,n}(\omega)\{\varphi_{n_1,n_2},\ \varphi_{N-n_1,N-n_2},\ -\varphi_{N-n_1,n_2},\ -\varphi_{n_1,N-n_2}\}^T\\ =&\bar{S}_{h,n}(\omega)\{\varphi_{n_1,n_2},\ \varphi_{N-n_1,N-n_2},\ -\varphi_{N-n_1,n_2},\ -\varphi_{n_1,N-n_2}\}^T,\end{aligned}$$

$S_{h,n}(\omega)$所对应的矩阵为

$$\bar{S}_{h;n}(\omega)=\begin{cases}\begin{pmatrix}1-\omega(\xi+\eta)&&&\\&1-\omega(2-\xi-\eta)&&\\&&1-\omega(1-\xi+\eta)&\\&&&1-\omega(1+\xi-\eta)\end{pmatrix}, & |n|<\dfrac{N}{2},\\ \begin{bmatrix}1-\omega(\xi+\eta)&\\&1-\omega(2-\xi-\eta)\end{bmatrix}, & n_1\text{或}n_2=\dfrac{N}{2},\\ [1-\omega(\xi+\eta)]_{1\times1}, & n_1=n_2=\dfrac{N}{2},\end{cases}\tag{4.12}$$

其中，$\xi=\sin^2\frac{n_1\pi}{2N}$，$\eta=\sin^2\frac{n_2\pi}{2N}$.

矩阵$\overline{S}_{h,n}(\omega)$中的元素是不难计算的.事实上，因$\lambda(S_{h,n}(\omega))=1-\frac{\omega}{4}\lambda(L_h)$，所以用（4.11）式中的诸量（$L_h$的特征值）代入，即可验证(4.12)式的正确性.例如，将$\lambda(L_h)=4\left(\cos^2\frac{n_1\pi}{2N}+\cos^2\frac{n_2\pi}{2N}\right)$代入，即有

$$\begin{aligned}\lambda(S_{h,n}(\omega))&=1-\frac{\omega}{4}\cdot 4\left(\cos^2\frac{n_1\pi}{2N}+\cos^2\frac{n_2\pi}{2N}\right)\\&=1-\omega(2-\xi-\eta).\end{aligned}$$

下面讨论I_h^{2h}所对应的特征值，由于

$$I_h^{2h}:\ E_{h,n}\rightarrow \operatorname{span}\{\Phi_{n_1,n_2}\},\quad |n|\leqslant\frac{N}{2}-1$$

仿前面办法，可求得全权限制算子I_h^{2h}所对应的变换矩阵

$$\overline{I}_h^{2h}=[(1-\xi)(1-\eta),\xi\eta,\xi(1-\eta),(1-\xi)\eta]^T. \tag{4.13}$$

同理易知I_{2h}^h：$\operatorname{span}\{\Phi_{n_1,n_2}\}\rightarrow E_{h,n}$的变换矩阵

$$\overline{I}_{2h}^h=[(1-\xi)(1-\eta),\xi\eta,\xi(1-\eta),(1-\xi)\eta]. \tag{4.14}$$

类似前面的作法，并利用(4.13)和(4.14)式，可求得

$$K_h^{2h}=I_h-I_{2h}^hL_{2h}^{-1}I_h^{2h}L_h$$

的变换矩阵是

$$\overline{K_{h,n}^{2h}}=\begin{cases}1-\frac{1}{d}[b_ic_j]_{4\times4}, & |n|<\frac{N}{2},\\ 2\times2\text{阶恒等阵}, & n_1\text{或}n_2=\frac{N}{2},\\ 1\times1\text{阶恒等阵}, & n_1=n_2=\frac{N}{2},\end{cases} \tag{4.15a}$$

其中，$d=\xi(1-\xi)+\eta(1-\eta)$，$b_1=(1-\xi)(1-\eta)$，$b_2=\xi\eta$，$b_3=\xi(1-\eta)$，$b_4=(1-\xi)\eta$，$c_1=b_1(\xi+\eta)$，$c_2=b_2(2-\xi-\eta)$，$c_3=b_3(1-\xi+\eta)$，$c_4=b_4(1+\xi-\eta)$.

最后，利用(4.8)、(4.12)和(4.15a)式，即可获得两层网格方法算子

$$M_h^{2h}:\ E_{h,n}\to E_{h,n}$$

的矩阵表示形式

$$\boldsymbol{M}_{h,n}^{2h}=\bar{S}_{h,n}^{\nu_2}(\omega)\bar{K}_{h,n}^{2h}\bar{S}_{h,n}^{\nu_1}(\omega). \qquad (4.15b)$$

这样，两层网格方法算子的谱半径 $\rho(M_h^{2h})$的计算问题，就转化为计算至多是4×4阶矩阵$\boldsymbol{M}_{h,n}^{2h}$的谱半径问题，即有

$$\rho(M_h^{2h})=\max\left\{|\lambda(\boldsymbol{M}_{h,n}^{2h})|\,\middle|\,|n|\leqslant\frac{N}{2}\right\}.$$

显然$\rho(M_h^{2h})$是依赖于参数 ω、h 及 $\nu=\nu_1+\nu_2$ 的（因为 $\rho(AB)=\rho(BA)$，所以M_h^{2h}的谱半径不单个依赖于 ν_1或ν_2），这样就可引进记号

$$\rho(h,\nu,\omega)=\rho(M_h^{2h}(h,\nu,\omega)),$$

$$\rho^*(\nu,\omega)=\sup\left\{\rho(h,\nu,\omega)\,\middle|\,h\leqslant\frac{1}{4}\right\},$$

其中 ρ 表示两层网格算子的收敛特性，它与 h 是有关的，但ρ^*表示它对一切细网格情形的收敛特性，是与 h 无关的。最后将计算出的ρ^*联同第三章（3.6）式计算出的光滑因子 $\mu^*(\omega)$一起列于表4.1中。

从表4.1中的数值可见如下几点事实：

(i) 当 $\omega=0.8$ 时，ω-Jacobi 方法的光滑效果和两层网格方法的收敛效果比$\omega=0.5$时相应的效果好，这正符合第三章§2.1中的分析。

(ii) 表中粗横线以上的数值表明 μ^*与 ρ^*等值；而表中横线以下的数值却反映了 μ^*与 ρ^* 的差值随着 ν 的增加而增加，即出现光滑因子减小、收敛因子增大的现象。这表明，光滑次数 ν 增大，虽然高频分量被大量衰减了，但经过粗网格法正后又引进了新的高频量，从而光滑作用受到了部分抵消。这一事实告诉我们，过多的光滑步是没有必要的，光滑效果应和粗网格校正的效果相匹配。表中横线以上的光滑步是理想的，例如，$\omega=0.8$，$\nu=2\sim3$较好。

表4.1　光滑因子$(\mu^*(\omega))^\nu$和两层两格方法收敛因子$\rho^*(\nu,\omega)$的比较

ν	$\omega=0.5$		$\omega=0.8$	
	$(\mu^*(\omega))^\nu$	$\rho^*(\nu,\omega)$	$(\mu^*(\omega))^\nu$	$\rho^*(\nu,\omega)$
1	0.750	0.750	0.600	0.600
2	0.563	0.563	0.360	0.360
3	0.422	0.422	0.216	0.216
4	0.316	0.316	0.130	0.137
5	0.237	0.237	0.078	0.113
6	0.178	0.178	0.047	0.079
7	0.133	0.133	0.028	0.085
8	0.100	0.118	0.017	0.076
9	0.075	0.106	0.010	0.068
10	0.056	0.097	0.006	0.062

(iii) 最令人兴奋的是ρ^*存在，即两层网格方法的收敛性与h无关。换句话说，对一切的细网格(例如h充分小)两层网格方法都有很好的收敛效果，这对于通常迭代敛速都与h有关来说是一个本质的进步。当然，两层网格方法的收敛因子$\rho(h,\nu,\omega)$严格地说是与h有关的，但随着网格加密，$\rho(h,\nu,\omega)$很快就趋于上确界$\rho^*(\nu,\omega)$，这一事实可以从表4.2的数值中清楚地看到。

本节就一个模型问题在有限维网格函数空间中分析了两层网格方法迭代矩阵$M_h^{2,h}$的特征值，从而定量地描述了两层网格方法的收敛特性，这种分析方法称为**模型分析方法**。用这种方法也可以分析使用其它光滑方法的模型问题两层网格方法的收敛特性[12-16]，但模型分析法太繁琐且适应性不好，在本章§4和§5中将用其它方法分析两层网格方法的h无关收敛性。

表4.2 随着网格加密$\rho(h,\nu,\omega)$很快趋于$\rho^*(\nu,\omega)$

h \ ρ \ ν \ ω	$\omega=0.5$				$\omega=0.8$			
	$\nu=1$	$\nu=2$	$\nu=3$	$\nu=4$	$\nu=1$	$\nu=2$	$\nu=3$	$\nu=4$
1/4	0.677	0.458	0.310	0.217	0.483	0.233	0.171	0.130
1/8	0.731	0.534	0.391	0.285	0.570	0.324	0.185	0.130
1/16	0.745	0.555	0.414	0.308	0.592	0.351	0.208	0.135
1/32	0.749	0.561	0.420	0.314	0.598	0.358	0.214	0.137
1/64	0.750	0.562	0.421	0.316	0.600	0.359	0.215	0.137
1/128	0.750	0.562	0.422	0.316	0.600	0.360	0.216	0.137
$\rho^*(\nu,\omega)$	0.750	0.563	0.422	0.316	0.600	0.360	0.216	0.137

§3 多层网格方法

3.1 多层网格方法V循环

两层网格方法中要求准确解出残差方程(2.3)，这仍是困难的而且是没有必要的，因在较粗网格层Ω_{2h}上解残差方程的工作量仍然相当大。人们完全有理由在Ω_{2h}网格层上不是解出(2.3)式中粗网格校正量的准确值，而只是用迭代法求出其近似值，且可以在更粗的网格层上重复使用粗网格校正过程和光滑过程，仅在最粗网格层上才准确求解残差方程。这种在多网格层上进行迭代校正的方法称为**多层网格方法**。由于在最粗网格层上求残差方程的工作量很小，内插、限制过程的工作量也很小，所以，若多层网格方法收敛，则它可以达到大幅度减少计算工作量、快速地数值求解微分方程的目的。

为此，定义一个**网格层序列**$\Omega_1,\Omega_2,\cdots,\Omega_l(l\geqslant 2)$，其中$\Omega_1$是确定的**最粗网格层**，$\Omega_l$ 是确定的**最细网格层**。相应地，$h_i(i=1,2,\cdots,l)$是Ω_i 层的网格步长参数；Ω_i 层上定义的网格函数空间记为 $G(\Omega_i)$；网格层上离散算子和迭代算子以及网格层间的限制和内插算子定义为

$$L_l:\ G(\Omega_l)\to G(\Omega_l),\qquad S_l:\ G(\Omega_l)\to G(\Omega_l),$$

$$I_l^{l-1}:\ G(\Omega_l)\to G(\Omega_{l-1}),\qquad I_{l-1}^{l}:\ G(\Omega_{l-1})\to G(\Omega_l);$$

ν_1^l表示在限制过程中于Ω_l网格层上所作的迭代次数，ν_2^l表示在内插过程中于Ω_l网格层上所作的迭代次数。

为了求解最细网格层上的离散方程

$$L_l u_l = f_l, \tag{4.16}$$

任意给定u_l的近似值u_l^j，经过多层网格方法的一个V循环后，可得到一个新的近似值u_l^{j+1}。在 l 个网格层上定义的**多层网格方法**的一个**V循环**步骤如下：

情形一，$l=2$，是两层网格方法，如§2所述。

情形二，$l>2$，是多层网格方法

1. 校正前光滑

对于任给的近似值u_l^j，对(4.16)式施行 $\nu_1^l(\geqslant 0)$光滑步后得$\bar{u}_l^j$，即

$$\bar{u}_l^j:=\mathrm{Relax}(u_l^j,L_l,f_l,\nu_1^l).$$

2. 粗网格校正

计算并限制残差 $d_{l-1}^j:=I_l^{l-1}\bar{d}_l^j=I_l^{l-1}(f_l-L_l\bar{u}_l^j)$；

计算校正量 以零为初值，对残差方程

$$L_{l-1}v_{l-1}^j=d_{l-1}^j$$

施行$l-1$层网格方法，求出近似校正量$\tilde{v}_{l-1}^j$；

内插校正量 $v_l^j:=I_{l-1}^l\tilde{v}_{l-1}^j$；

计算被校正的近似值 $\bar{\bar{u}}_l^j:=\bar{u}_l^j+v_l^j$。

3. 校正后光滑

以$\bar{\bar{u}}_l^j=\bar{u}_l^j+v_l^j$为初值，在$\Omega_l$层上对(4.16)式施行 ν_2^l步迭代，

便得到经过一个多层网格V循环迭代步后的新近似值，即

$$u_l^{j+1} = \mathrm{Relax}(\tilde{u}_l^j, L_l, f_l, \nu_2^l).$$

为了进一步加深对多层网格方法实现过程的理解，下面作出3层网格上的一个V循环的具体计算步骤解析（参见图4.2）和l个网格层上利用V循环求解离散方程（4.16）的计算框图（参见图4.3）。

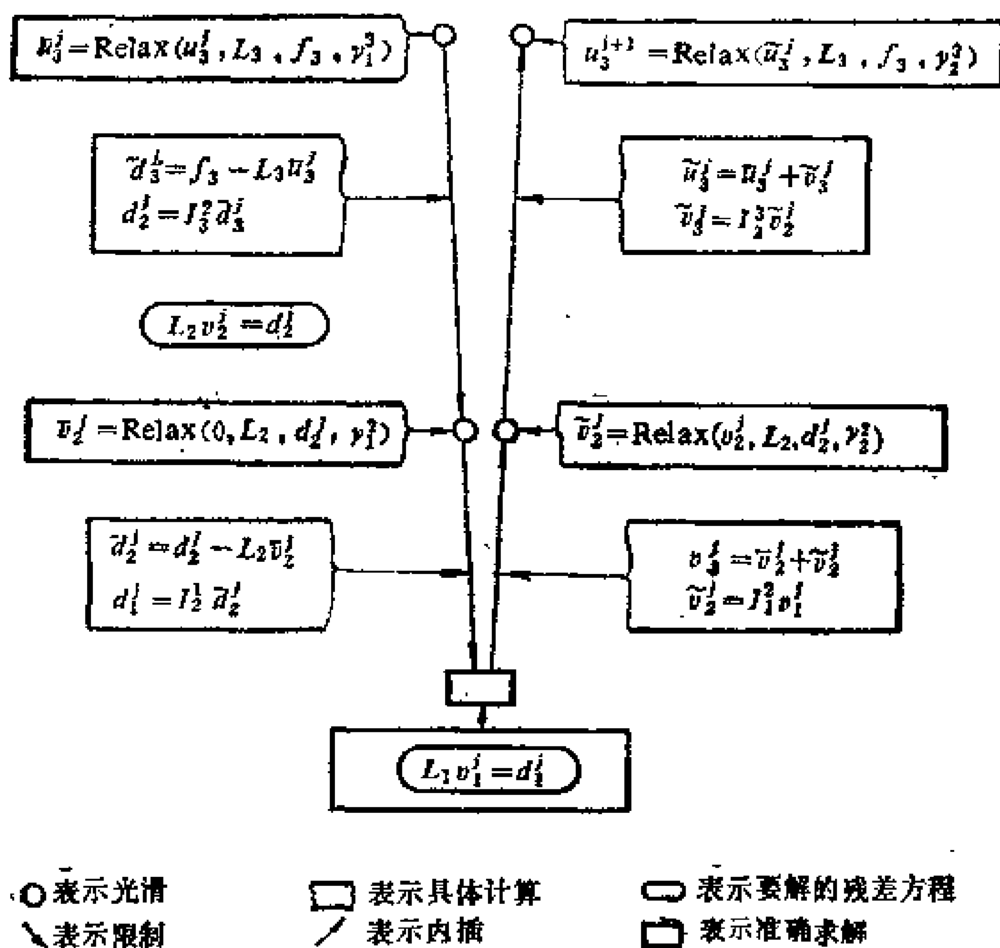


图4.2　3层网格上的一个V循环迭代步解析

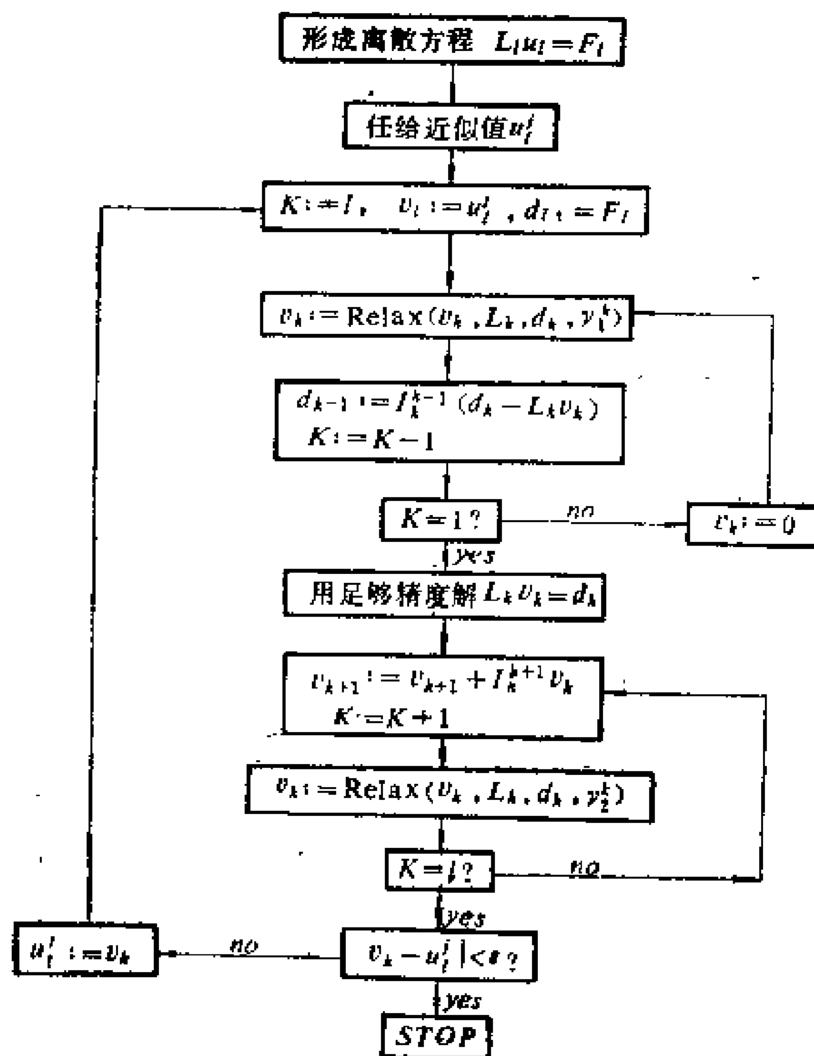


图4.3 l层网格V循环解离散方程的FORTRAN流程图

由图 4.2 展示V循环的计算步骤解析可见，在每个网格层上都进行了如下几个步骤：校正前迭代，限制，找残差方程的较好的近似解，内插，校正后迭代。它类同于两层网格方法。因此，多层网格方法V循环实际上是以两层网格方法为基础的逐步递推结构，参见图4.4。

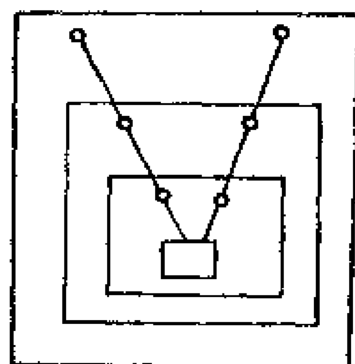


图 4.4

3.2 其它循环方式的多层网格方法

l 层网格V循环方法实际上是定义在Ω_l和Ω_{l-1}层上的两层网格方法，只不过应用了$l-1$层网格V循环的一个迭代步去近似求解Ω_{l-1}网格层上的残差方程 $L_{l-1}v_{l-1}=d_{l-1}$。

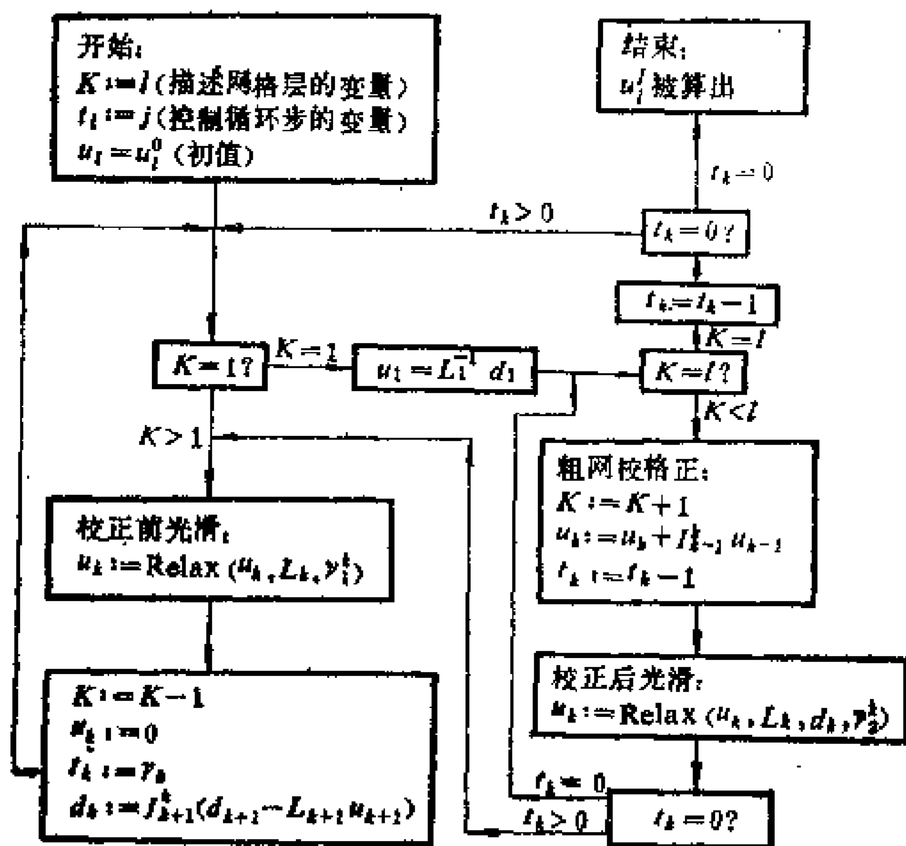


图 4.5

同样地，可以应用$l-1$层网格V循环的γ个迭代步或其它方式，达到近似求解Ω_{l-1}网格层上的残差方程的目的，这样就得到了种种不同循环方式的多层网格方法，图 4.5 表示了它们的 FORTRAN流程。图 4.5 只描述了j个循环步，它是由t_l控制的；γ的不同数值可以实现不同形式的循环过程。当$\gamma=1$时，它是**V循环**，图4.6*a*）简单地表示了它的一个循环步；$\gamma=2$时，它是**W循环**，图4.6*b*）、*c*）简示了它的一个循环步的结构；$\gamma=1$(k为偶数）$\gamma=2$(k为奇数)时，它的循环结构如图 4.6*d*）、*e*）、*f*）所示；也可以设置其它的γ_k，设计出不同结构的多层网格方法。

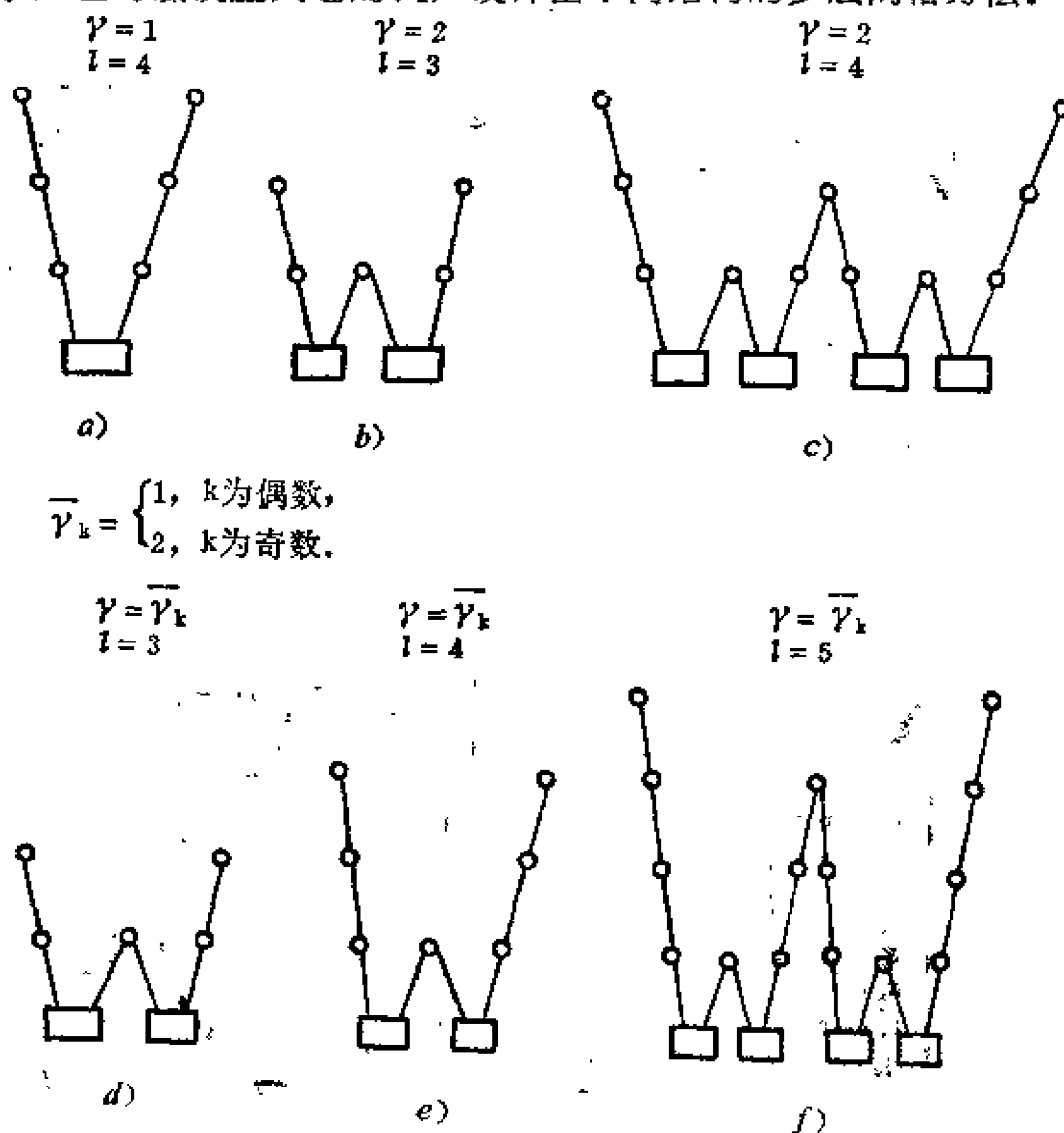


图 4.6

多层网格方法的计算量为$O(N_l)$(参见本章§3.4，其中N_l为Ω_l的节点数)，这是一个很好的指标。多层网格方法所需的存贮量也很小，除必要时需存贮系数矩阵非零元素外，一般只需$2\times\frac{4}{3}$个数组，图 4.7 表示多层网格方法中存贮单元的使用情况，这里仅以V循环为例，图中↘表示形成、限制并存贮残差量，↑表示松弛迭代形成并存贮校正量，↶表示校正存贮过程。

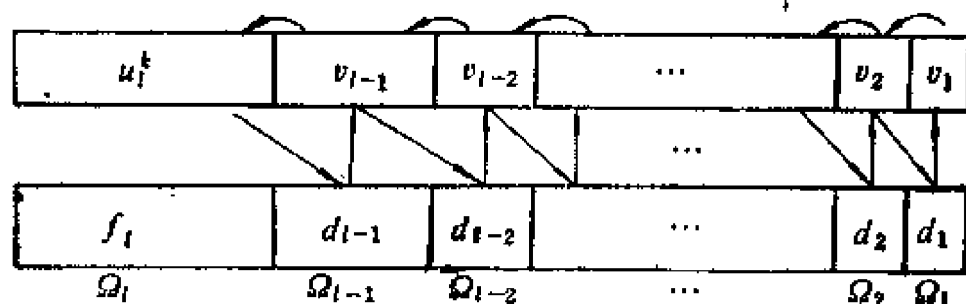


图 4.7

多层网格方法的收敛效果也是好的且与h无关，收敛性分析可见本章§4，这里只举出一个数例。在单位正方形区域上，采用5点离散格式解Poisson 方程，按图4.5循环方式计算，选用的参数是$\gamma_1^k=2$，$\gamma_2^k=0$，$l=5$，则有

当$\gamma=1$时，多层网格方法的收敛因子是0.1682，

当$\gamma=2$时，多层网格方法的收敛因子是0.0736，

当$\gamma=3$时，多层网格方法的收敛因子是0.0735，

通常，W循环($\gamma=2$)的收敛率略高于V循环($\gamma=1$)的收敛率，但W循环却多花费了一些计算量。$\gamma\geqslant3$时的收敛率并不比$\gamma=2$时好多少，它们并不适用。$\gamma=1$(k 为偶数或奇数)、$\gamma=2$(k 为奇数或偶数)的循环方式在半粗化情形和红黑粗化情形也是有用的。

由于迭代效果与初始值的选取密切相关，所以改善初值后再运用多层网格方法对减少整体计算量是大有好处的，但不会改善多层网格方法的收敛率。人们可采用下面两种办法改善初值。办法一，在最粗网格层Ω_1上解出数值解，再将其一层一层地内插，

直至最细网格层Ω_l，这样可得到一个较好的初值。办法二，在最粗网格层Ω_1上解出数值解，再将其内插到 Ω_2 上且作适当次数迭代，不断重复内插和迭代这两个步骤，直至在最细网格层Ω_l上获得初值为止。用办法二获得的初值当然比用办法一获得的初值要好些。

3.3 完全多层网格方法

根据改善迭代初值的考虑，人们提出了不断改善初值不断进行多层网格循环的完全多层网格方法。其基本实现过程如下：先在Ω_1网格层上解$L_1u_1=f_1$；将解出的u_1内插到 Ω_2网格层，从而得到用于求解 $L_2u_2=f_2$ 的一个较好的初值；利用多层网格方法的适当个数的迭代步（使用循环方式参数 γ）求得u_2；再不断使用内插和多层网格循环过程，直到在Ω_l层上求出离散方程 $L_lu_l=f_l$ 满足精度要求的近似解。

设在Ω_k网格层上，按图4.5的多层网格迭代步去寻求 $L_ku_k=f_k$ 的近似解（初值是 u_k^0），此多层网格方法记为

$MG(u_k^0,L_k,f_k,\gamma_k)$：$G(\Omega_k)\to G(\Omega_k)$，

则求解(4.16)式的**完全多层网格方法**的实现过程可由图4.8描述。

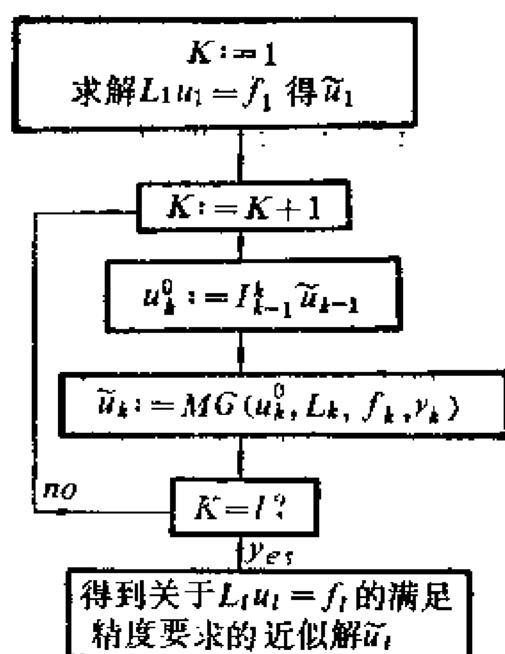


图 4.8

完全多层网格算法利用了**嵌套迭代**这个计算原则，它的效率不仅依赖于作为基础的多层网格方法的效率；还依赖于层间数值转换的控制，在Ω_k层上让$\tilde{u}_k$的精度过高是会付出很大代价的而且是没有必要的，一般让$\tilde{u}_k$满足离散方程$L_ku_k=f_k$所要求的数值解精度即可。另外，在完全多层网

表 4.3　采用线性内插时，完全多层网格方法的计算结果

h_k	$\nu=\gamma=2$; $i=1$	整体误差 $\Vert\tilde{u}_k^i-u\Vert_0$	迭代误差 $\Vert\tilde{u}_k^i-u_k\Vert_0$	离散误差 $\Vert u_k-u\Vert_0$	$\nu=\gamma=2$; $i=2$	整体误差 $\Vert\tilde{u}_k^i-u\Vert_0$	迭代误差 $\Vert\tilde{u}_k-u_k\Vert_0$
1/2	$\tilde{u}_0$	$1.45E-4$	0	$1.45E-4$	$\tilde{u}_0$	$1.45E-4$	0
1/4	$\tilde{u}_1^0$	$2.23E-2$	$2.23E-2$		$\tilde{u}_1^0$	$2.23E-2$	$2.23E-2$
	$\tilde{u}_1^1$	$1.57E-3$	$1.49E-3$	$8.81E-5$	$\tilde{u}_1^1$	$1.57E-3$	$1.49E-3$
					$\tilde{u}_1^2$	$1.72E-4$	$9.16E-5$
1/8	$\tilde{u}_2^0$	$7.26E-3$	$7.24E-3$		$\tilde{u}_2^0$	$6.49E-3$	$6.47E-3$
	$\tilde{u}_2^1$	$6.19E-4$	$5.98E-4$	$2.64E-5$	$\tilde{u}_2^1$	$5.13E-4$	$4.92E-4$
					$\tilde{u}_2^2$	$5.23E-5$	$2.86E-5$
1/16	$\tilde{u}_3^0$	$2.10E-3$	$2.09E-3$		$\tilde{u}_3^0$	$1.73E-3$	$1.73E-3$
	$\tilde{u}_3^1$	$1.34E-4$	$1.30E-4$	$6.89E-6$	$\tilde{u}_3^1$	$1.03E-4$	$9.91E-5$
					$\tilde{u}_3^2$	$9.89E-6$	$4.91E-6$
1/32	$\tilde{u}_4^0$	$5.18E-4$	$5.17E-4$		$\tilde{u}_4^0$	$4.45E-4$	$4.43E-4$
	$\tilde{u}_4^1$	$2.59E-5$	$2.54E-5$	$1.74E-6$	$\tilde{u}_4^1$	$1.87E-5$	$1.82E-5$
					$\tilde{u}_4^2$	$2.02E-6$	$8.52E-7$
1/64	$\tilde{u}_5^0$	$1.24E-4$	$1.23E-4$		$\tilde{u}_5^0$	$1.12E-4$	$1.12E-4$
	$\tilde{u}_5^1$	$4.81E-6$	$4.76E-6$	$4.36E-7$	$\tilde{u}_5^1$	$3.32E-6$	$3.25E-6$
					$\tilde{u}_5^2$	$4.63E-7$	$1.47E-7$

表4.4 采用三次插值时，完全多层网格方法的计算结果

h_k	$\nu=\gamma=2;\ i=1$				$\nu=\gamma=2;\ i=2$		
		整体误差 $\Vert\tilde{u}_k^j-u\Vert_0$	迭代误差 $\Vert\tilde{u}_k^j-u_k\Vert_0$	离散误差 $\Vert u_k-u\Vert_0$		整体误差 $\Vert\tilde{u}_k^j-u\Vert_0$	迭代误差 $\Vert\tilde{u}_k^j-u_k\Vert_0$
1/2	$\tilde{u}_0$	1.45E−4	0	1.45E−4	$\tilde{u}_0$	1.45E−4	0
1/4	$\tilde{u}_1^0$	9.85E−3	9.81E−3		$\tilde{u}_1^0$	9.85E−3	9.81E−3
	$\tilde{u}_1^1$	3.06E−4	2.59E−4	8.81E−5	$\tilde{u}_1^1$	3.06E−4	2.59E−4
					$\tilde{u}_1^2$	9.04E−5	4.14E−5
1/8	$\tilde{u}_2^0$	1.15E−3	1.14E−3		$\tilde{u}_2^0$	9.98E−4	9.89E−4
	$\tilde{u}_2^1$	1.21E−4	1.08E−4	2.64E−5	$\tilde{u}_2^1$	9.34E−5	8.00E−5
					$\tilde{u}_2^2$	2.86E−5	4.22E−6
1/16	$\tilde{u}_3^0$	1.75E−4	1.72E−4		$\tilde{u}_3^0$	1.00E−4	9.77E−5
	$\tilde{u}_3^1$	2.00E−5	1.70E−5	6.89E−6	$\tilde{u}_3^1$	1.22E−5	8.71E−6
					$\tilde{u}_3^2$	6.98E−6	4.65E−7
1/32	$\tilde{u}_4^0$	2.37E−5	2.29E−5		$\tilde{u}_4^0$	1.13E−5	1.02E−5
	$\tilde{u}_4^1$	2.93E−6	2.19E−6	1.74E−6	$\tilde{u}_4^1$	2.01E−6	8.34E−7
					$\tilde{u}_4^2$	1.74E−6	4.14E−8
1/64	$\tilde{u}_5^0$	3.16E−6	2.91E−6		$\tilde{u}_5^0$	1.91E−6	1.52E−6
	$\tilde{u}_5^1$	5.14E−7	2.59E−7	4.36E−7	$\tilde{u}_5^1$	4.47E−7	7.62E−8
					$\tilde{u}_5^2$	4.30E−7	4.03E−8

格方法中应要求插值(不是MG循环步内的插值)误差阶至少应等于相对相容误差阶。

特别地，当采用高阶插值时，在每个网格层上只要作一个MG循环步即可使完全多层网格方法顺利进行[18]。

下面列出用完全多层网格方法数值求解

$$\begin{cases} -\Delta u = -3y/(x+1) - y^3/(x+1)^3, & (x,y)\in(0,1)\times(0,1), \\ u = 0.5y^3/(x+1), & (x,y)\in\partial\Omega \end{cases}$$

的数据。此边值问题的真解是$u=0.5y^3/(x+1)$，用u_k表示此问题在Ω_k层上的准确的数值解，$\tilde{u}_k$表示在Ω_k层上用完全多层网格方法获得的数值解，$\|\cdot\|$表示Euclid 模。在计算中要用到带有循环参数$\gamma=2$的W循环，层间光滑次数一律取$\nu=2$，多层网格循环步数用i表示；表4.3中列出了采用线性内插的数值结果，容易看出，每层作2个MG循环步比作1个MG循环步要好一点；表4.4中列出了采用三次插值的数值结果，容易看出，一个MG循环步就能取得满意的效果。

3.4 计算工作量估计

设W_1表示在最粗网格层Ω_1上解残差方程所需的计算量；W_{k+1}^k表示在Ω_{k+1}层上的光滑、Ω_{k+1}与Ω_k层间限制和内插的计算量；W_k表示在$\Omega_k,\cdots,\Omega_1$网格层上进行k层网格方法的总计算工作量；利用多层网格方法的递推定义，立即可得其整体计算量的估计符合如下递推关系

$$\begin{cases} W_2 = W_2^1 + W_1, \\ W_{k+1} = W_{k+1}^k + \gamma_k W_k, \quad k=2,\cdots,l-1. \end{cases} \tag{4.17}$$

如果γ_k与k无关，那么，由(4.17)式得

$$W_l = \sum_{k=2}^{l} \gamma^{l-k} W_k^{k-1} + \gamma^{l-1} W_1, \quad l\geqslant 2. \tag{4.18}$$

例如，对标准粗化情形，粗细网格点数目之间的关系是$N_k \approx 4N_{k-1}$，$k=2,3,\cdots,l$。由于采用局部格式，在Ω_k层上作少数

几个光滑步的工作量是$O(N_k)$；Ω_k 和 Ω_{k-1} 层间的限制和内插计算量也是$O(N_k)$；所以$W_k^{k-1}=O(N_k)$。因此，利用(4.18)式，可知一个V循环($\gamma_k\equiv 1$)的总计算量为

$$
\begin{aligned}
W_l &= W_l^{l-1}+W_{l-1}^{l-2}+\cdots+W_2^{1}+W_1\\
&\approx C\cdot(N_l+N_{l-1}+\cdots+N_2+N_1)\\
&\approx CN_l\left(1+\frac{1}{4}+\frac{1}{16}+\cdots+\frac{1}{4^{l-1}}\right)\approx\frac{4}{3}CN_l.
\end{aligned}
$$

相同的道理可得

$$
W_l\approx\begin{cases}\dfrac{4}{3}CN_l, & \gamma=1,\\ 2CN_l, & \gamma=2,\\ 4CN_l, & \gamma=3,\\ O(N_l\lg N_l), & \gamma=4.\end{cases}\tag{4.19}
$$

对标准粗化或其它网格分划情形，当γ_k恒为常数且$N_k\approx\tau N_{k-1}$时，利用(4.18)式可得

$$
W_l\approx\begin{cases}\dfrac{\tau}{\tau-\gamma}CN_l, & \gamma<\tau,\\ O(N_l\lg N_l), & \gamma=\tau.\end{cases}\tag{4.20}
$$

(4.20)式表明，当$\gamma<\tau$时多层网格方法的一个循环步所需的计算量与最细网格层节点数成正比，因只需有限个多层网格方法循环步，所以用多层网格方法求解细网格层上离散方程的总计算量仍与最细网格层节点数成正比，即

$$
W=O(N_l).
$$

这一指标从根本上克服了经典迭代法解(4.16)式的工作量一般为$O(N_l^{3/2})\sim O(N_l^{2})$这个缺点。下面§4还证明了多层网格方法的敛速与$h$无关，这就是说，无论分划$\Omega_l$多么细密，多层网格方法都能保持恒定的敛速，且在有限步内完成求解工作，这一特性又从根本上克服了经典迭代法的敛速与h有关，Ω_l越细所需迭代步越多，敛速越慢这个缺点。所以$\gamma<\tau$时的多层网格方法是渐近最优

的算法。另外，多层网格方法仍保持了迭代法的种种优点，因此作为迭代法，多层网格方法取得各项指标的全面优势而受到人们的重视。

对于完全多层网格方法，很容易估计其整体计算量W，例如对标准粗化情形有保守估计：

$$W=\frac{4}{3}\gamma W_l+\frac{4}{3}W_{l-1}^{INT}, \tag{4.21}$$

其中，W_l如(4.18)式所示，W_{l-1}^{INT}表示由Ω_{l-1}到Ω_l插值过程所需的工作量，γ表示循环方式参数或表示层间所作的多层网格循环步数，一般地，完全多层网格方法所需的计算量比一般多层网格方法所需计算量少些。完全多层网格方法也是一种渐近最优算法。

§4 多层网格方法的h无关收敛性

4.1 多层网格方法迭代算子及其收敛性

设$\gamma=\gamma_k$，M_l表示在l层网格上施行多层网格方法的迭代算子，利用多层网格方法的递推定义有下面定理。

定理4.1 M_l适合递归关系：

$$\begin{cases}M_2=S_2^{\nu_2}(I_2-I_1^2L_l^{-1}I_2^1L_2)S_2^{\nu_1},\\ M_{k+1}=S_{k+1}^{\nu_2}(I_{k+1}-I_k^{k+1}(I_k-(M_k)^{\gamma})L_k^{-1}I_{k+1}^kL_{k+1})S_{k+1}^{\nu_1},\\ k=2,3,\cdots,l-1.\end{cases} \tag{4.22}$$

证明 由(4.8)式即知M_2是正确的。

若在Ω_{k+1}和Ω_k层上施行两层网格方法，则据(4.8)式有

$$M_{k+1}=S_{k+1}^{\nu_2}(I_{k+1}-I_k^{k+1}L_k^{-1}I_{k+1}^kL_{k+1})S_{k+1}^{\nu_1}. \tag{4.23}$$

然而在多层网格方法中，并不是直接解出Ω_k层上的残差方程，而是在$\Omega_k,\cdots,\Omega_1$诸 网格层上，使用γ个完整的k层多层网格方法迭代步(循环方式与γ有关，且初值为零)求Ω_k层上残差方程的近似值，这相当于求解残差方程

$$L_k w_k = d_k$$

的近似值，使用了迭代算子为 M_k（k层多层网格方法迭代算子）的迭代方法：

$$\begin{cases} w_k^{l+1} = M_k w_k^l + B, \\ w_k^0 = 0, \end{cases}$$

并迭代了 γ 步（注意，γ 是循环参数，参见图4.6）。所以有

$$\begin{aligned} w_k^\gamma &= M_k w_k^{\gamma-1} + B = M_k(M_k w_k^{\gamma-2} + B) + B \\ &= (I_k + M_k + M_k^2 + \cdots + M_k^{\gamma-1} + M_k^\gamma w_k^0)B \\ &= (I_k - (M_k)^\gamma)B/(I_k - M_k) \\ &= (I_k - (M_k)^\gamma)w_k = (I_k - (M_k)^\gamma)L_k^{-1}d_k, \end{aligned}$$

故只需用 $\Omega_k, \cdots, \Omega_1$ 网格层上的多层网格迭代（初值为零，循环参数为γ）算子 $(I_k - (M_k)^\gamma)L_k^{-1}$ 取代(4.23)式中的 L_k^{-1} 即得(4.22)式中的 M_{k+1}。 □

为了证明收敛性时方便，将(4.22)式改写为

$$\begin{cases} M_{k+1} = M_{k+1}^k + A_k^{k+1}(M_k)^\gamma A_{k+1}^k, \quad k = 2,3,\cdots,l-1, \\ A_k^{k+1} = S_{k+1}^{\nu_2} I_k^{k+1}, \\ A_{k+1}^k = L_k^{-1} I_{k+1}^k L_{k+1} S_{k+1}^{\nu_1}, \\ M_{k+1}^k = S_{k+1}^{\nu_2}(I_{k+1} - I_k^{k+1} L_k^{-1} I_{k+1}^k L_{k+1}) S_{k+1}^{\nu_1}. \end{cases} \tag{4.24}$$

设 $\|\cdot\|$ 代表任何一种适当的模，则利用 (4.24) 式可得多层网格方法的收敛性定理。

定理4.2 若对 $\forall\, k \leqslant l-1$，一致有

$$\|M_{k+1}^k\| \leqslant \sigma^*, \quad \|A_k^{k+1}\| \cdot \|A_{k+1}^k\| \leqslant C,$$

成立，则有 $\|M_l\| \leqslant \eta_l$，且 η_l 合递推关系

$$\eta_1 = \sigma^*, \quad \eta_{k+1} = \sigma^* + C(\eta_k)^\gamma \quad (k = 1,2,\cdots,l-1).$$

应该看到，由于光滑作用的 h 无关性，内插限制算子的 h 无关有界性，所以 $\|A_k^{k+1}\| \cdot \|A_{k+1}^k\|$ 的 h 无关有界性是容易满足的；因此，多层网格方法的 h 无关收敛性归结为两层网格方法的 h 无关收敛性。

4.2 两层网格方法的h无关收敛性

这里考虑确定在Ω_{k-1}和Ω_k上的两层网格方法，其迭代算子为

$$M_k^{k-1}=S_k^{\nu_2}(I_k-I_{k-1}^kL_{k-1}^{-1}I_k^{k-1}L_k)S_k^{\nu_1},\ \ k\geqslant 2.$$

为了下面证明时方便，设$\nu_2=0$，$\nu_1=\nu$，即只有校正前光滑不作校正后光滑，那么两层网格方法迭代算子可改写为

$$\begin{aligned}M_k^{k-1}&=(I_k-I_{k-1}^kL_{k-1}^{-1}I_k^{k-1}L_k)S_k^\nu\\&=(L_k^{-1}-I_{k-1}^kL_{k-1}^{-1}I_k^{k-1})L_kS_k^\nu,\quad k\geqslant 2.\end{aligned}$$

因此，关键是估计$\|L_kS_k^\nu\|$和$\|L_k^{-1}-I_{k-1}^kL_{k-1}^{-1}I_k^{k-1}\|$。

容易理解，$\|L_kS_k^\nu\|$和$\|S_k^\nu\|$都描述了松弛迭代的光滑效果，因为$\|S_k^\nu v_k\|/\|v_k\|$与$\|L_kS_k^\nu v_k\|/\|L_k\|\|v_k\|$一样反映了松弛迭代的光滑特征且$\sup\|L_kS_k^\nu\|/\|L_k\|$可作为光滑因子，所以$\|L_kS_k^\nu\|$可用以反映两层网格方法中的光滑迭代步特征。另一方面，$\|L_k^{-1}-I_{k-1}^kL_{k-1}^{-1}I_k^{k-1}\|$描述了粗网格层的解$u_{k-1}\in G(\Omega_{k-1})$经限制和内插算子作用后近似细网格层上解$u_k\in G(\Omega_k)$的效果，所以$\|L_k^{-1}-I_{k-1}^kL_{k-1}^{-1}I_k^{k-1}\|$可用以反应两层网格方法中的粗网格校正步的特征。这样，可以利用这两种模量来估计两层网格方法的收敛性。

定理4.3 如果

"光滑特性"

$$\begin{cases}\|L_kS_k^\nu\|\leqslant\eta(\nu)h_k^{-\alpha},\ \ k\geqslant 2,\\ \eta(\nu)\to 0,\quad \nu\to\infty,\quad \eta(\nu)\text{与}h\text{无关};\end{cases}\tag{4.25}$$

和**"逼近特性"**

$$\|L_k^{-1}-I_{k-1}^kL_{k-1}^{-1}I_k^{k-1}\|\leqslant ch_k^\alpha,\ \ k\geqslant 2,\tag{4.26}$$

c是与h无关的常数；

满足，则存在$\bar\nu$，使有

$$\|M_k^{k-1}\|\leqslant c\eta(\bar\nu)\leqslant\rho<1.$$

定理4.3的证明是显然的，不必赘述。值得指出的是，利用"光滑特性"和"逼近特性"证明两层网格方法收敛的办法首先由

W. Hackbusch引用的[13]，这一套办法由于采用了模的估计，其光滑特性的描述不同于第三章§2；两层网格方法收敛性证明和描述办法也不同于模型分析法。模型分析法虽能定量分析两层网格方法的收敛率，但局限性大且过于烦杂。而定理 4.3 的分析办法虽只能定性地描述两层网格方法的收敛率，但它的适应性较广.下面将分析差分算子的“逼近特性”和迭代算子的“光滑特性”，由于本篇内容并不以完美无缺的复杂证明为主，而只是以应用和掌握研究问题的基本办法为主，故本篇中对差分算子和有限元算子的“逼近特性”作了详细的证明，但对于迭代算子的“光滑特性”，只是就模型问题的几个简单迭代算子作了分析，其它情形读者可参阅[13]，或自行推导证明。

4.3 关于差分算子“逼近特性”的估计

定理4.4 如果下列条件适合：

$$\|L_k^{-1}\|\leqslant c,\ \|L_k\|\leqslant c\quad (k\geqslant 1), \tag{4.27}$$

$$\|I_k^{k-1}\|\leqslant c,\ \|\tilde{I}_k^{k-1}\|\leqslant c\quad (k\geqslant 2), \tag{4.28}$$

$$\|\sigma_{k-1}\|=\|I_k^{k-1}L_kI_{k-1}^k-L_{k-1}\|\leqslant ch_{k-1}^{\alpha}\quad (k\geqslant 2) \tag{4.29}$$

$$\|I-I_{k-1}^k\tilde{I}_k^{k-1}\|\leqslant ch_{k-1}^{\alpha}\quad (k\geqslant 2), \tag{4.30}$$

$$\frac{1}{c}\|v_{k-1}\|\leqslant\|I_{k-1}^kv_{k-1}\|\leqslant c\|v_{k-1}\|\quad (\forall v_{k-1}\in G_{k-1},k\geqslant 2), \tag{4.31}$$

则“逼近特性”

$$\|L_k^{-1}-I_{k-1}^kL_{k-1}^{-1}I_k^{k-1}\|\leqslant ch_{k-1}^{\alpha}\quad (\forall k\geqslant 2) \tag{4.32}$$

成立。

证明 任意引进一个适合(4.28)式要求的限制算子$\tilde{I}_k^{k-1}$，并将(4.32)式左端变为

$$\begin{aligned}
&\|L_k^{-1}-I_{k-1}^kL_{k-1}^{-1}I_k^{k-1}\|=\|(I_k-I_{k-1}^kL_{k-1}^{-1}I_k^{k-1}L_k)L_k^{-1}\|\\
&\quad=\|(I_k-I_{k-1}^kL_{k-1}^{-1}I_{k-1}^kL_k)(I_k-I_{k-1}^k\tilde{I}_k^{k-1})L_k^{-1}\\
&\quad-I_{k-1}^kL_{k-1}^{-1}(I_k^{k-1}L_kI_{k-1}^k-L_{k-1})\tilde{I}_k^{k-1}L_k^{-1}\|\\
&\quad\leqslant\{1+\|I_{k-1}^k\|\cdot\|L_{k-1}^{-1}I_k^{k-1}L_k\|\}\cdot\|(I_k-I_{k-1}^k\tilde{I}_k^{k-1})L_k^{-1}\|
\end{aligned}$$

$$+\|I_{k-1}^k\|\cdot\|L_{k-1}^{-1}\delta_{k-1}\tilde{I}_k^{k-1}L_k^{-1}\|.$$

再分别估计上式中出现的几个模：由(4.31)式有

$$\|I_{k-1}^k\|\leqslant c,$$

由(4.27)和(4.28)式有

$$\|L_{k-1}^{-1}I_k^{k-1}L_k\|\leqslant c,$$

由(4.27)、(4.28)和(4.30)式有

$$\|(I_k-I_{k-1}^k\tilde{I}_k^{k-1})L_k^{-1}\|\leqslant ch_{k-1}^{\alpha},$$

由(4.27)～(4.29)式有

$$\|L_{k-1}^{-1}\delta_{k-1}\tilde{I}_k^{k-1}L_k^{-1}\|\leqslant ch_{k-1}^{\alpha}.$$

综合上述诸式，证得(4.32)式成立。 □

读者不难体会定理4.4中诸条件的意义。(4.27)和(4.28)式分别描述了离散算子L_k的正则性[4,5]和限制算子I_k^{k-1}、$\tilde{I}_k^{k-1}$的有界性，$\tilde{I}_k^{k-1}$是证明过程中引进的一个有界限制算子；对$\tilde{I}_k^{k-1}$的进一步要求表现于(4.30)式中，这相当于要求$\tilde{I}_k^{k-1}$与I_{k-1}^k的近似互逆程度；(4.31)式要求内插算子的一致有界性；在(4.27)～(4.31)式的诸条件中，最关键的条件是L_k的正则性和(4.29)式，后者描述了算子L_k经内插和限制算子作用后与L_{k-1}算子的逼近程度，这个条件相当于要求内插算子I_{k-1}^k和限制算子I_k^{k-1}的近似互逆程度，它反映了粗细网格层上的网格函数在转化过程中的正当要求。

4.4 关于迭代算子"光滑特性"的估计

这里以Helmholtz方程$-\Delta u+cu=f$为例($c=0$，则简化为Poisson方程)，定解区域$\Omega\subset R^2$被矩形网格离散后作坐标式编号；Ω_k层上的5点离散格式记为L_k；光滑迭代采用GS方法，其迭代矩阵记为S_k，显然

$$
\begin{cases}
L_k=\begin{pmatrix} & -1 & \\ -1 & 4+ch_k^2 & -1 \\ & -1 & \end{pmatrix},\ L_k^d=\begin{pmatrix} & 0 & \\ 0 & 4+ch_k^2 & 0 \\ & 0 & \end{pmatrix}, \\
L_k^l=\begin{pmatrix} & 0 & \\ 1 & 0 & 0 \\ & 1 & \end{pmatrix}, \qquad L_k^r=\begin{pmatrix} & 1 & \\ 0 & 0 & 1 \\ & 0 & \end{pmatrix}, \\
L_k=L_k^d-L_k^l-L_k^r, \qquad S_k=(L_k^d-L_k^l)^{-1}L_k^r.
\end{cases} \tag{4.33}
$$

这种情形的GS方法是经常碰到的，它的“光滑特性”也容易分析。

定理4.5 设$L_k=L_k^d-L_k^l-L_k^r$对称，$L_k^d=d_kI_k$，$d_k>0$，$\|L_k^l\|=\|L_k^r\|\leqslant d_k/2$，那么

$$
\|L_kS_k^\nu\|\leqslant d_k\cdot\begin{cases} 2/3^\nu, & \nu=1,2, \\ \dfrac{1}{2}(\nu-2)^{\frac{\nu-2}{2}}\Big/\nu^{\frac{\nu}{2}}, & \nu\geqslant 3. \end{cases} \tag{4.34}
$$

证明 因为

$$
L_kS_k^\nu=(d_kI_k-L_k^l-L_k^r)(d_kI_k-L_k^l)^{-\nu}(L_k^r)^\nu,
$$

所以

$$
\rho(L_kS_k^\nu)=\|L_kS_k^\nu\|=\max\{(d_k-2\mathrm{Re}\lambda)|\lambda/(d_k-\lambda)|^\nu\},
$$

其中λ是L_k^l和L_k^r的特征值，$\mathrm{Re}\lambda$是λ的实部。

令$x=\mathrm{Re}\lambda$，$y=|\lambda|$，则$x\in[-y,+y]$，$y\in[0,d_k/2]$，从而上式可改写为

$$
\begin{aligned}
\|L_kS_k^\nu\|&=\Phi(\lambda)=\Phi(x,y) \\
&=(d_k-2x)y^\nu(d_k^2-2xd_k+y^2)^{-\nu/2} \\
&=(d_k-2x)\Big/\left(1+\frac{d_k^2-2xd_k}{y^2}\right)^{\nu/2},
\end{aligned}
$$

显然

$$
\Phi(x,y)\leqslant\Phi(x,d_k/2)=d_k\left(1-\frac{2x}{d_k}\right)\Big/\left(5-\frac{8x}{d_k}\right)^{\nu/2}.
$$

易知当$\nu\geqslant3$时，$\Phi(x,d_k/2)$在$x=d_k\left(\frac{1}{2}-1/4(\nu-2)\right)$处取得极大值$\frac{d_k}{2}(\nu-2)^{\frac{\nu-2}{\nu}}\Big/\nu^{\frac{\nu}{2}}$；当$\nu=1,2$时，因$-d_k/2\leqslant x\leqslant d_k/2$，

所以$\Phi(x,d_k/2)$在$x=-d_k/2$处取得极大值$2d_k/3^\nu$，故证得(4.34)式。 □

(4.34)式右端是一个关于ν的函数，可记为$\varphi(\nu)$，当然可把它改写成(4.25)式右端项的形式。事实上，$\varphi(\nu)\cdot h_k^\alpha\cdot h_k^{-\alpha}\leqslant\eta(\nu)\cdot h_k^{-\alpha}$，$\eta(\nu)$与$h$无关，且当$\nu\to+\infty$时$\eta(\nu)\to 0$。由此可知，本例中的GS迭代是满足"光滑特性"(4.25)式的。

§5 有限元多层网格方法

5.1 有限元多层网格方法的代数解释

考虑变分问题

$$J(u)=a(u,u)-2(f,u)\Rightarrow\min,\quad\forall u\in H_E^1(\Omega),\tag{4.35}$$

及其相应的变分方程

$$a(u,v)=(f,v),\quad u\in H_E^1(\Omega),\quad\forall v\in H_0^1(\Omega),\tag{4.36}$$

其中，函数类$H_E^1(\Omega)$和Sobolev空间$H_0^1(\Omega)$的意义参见第一篇第二章§1，这里还假设(4.35)或(4.36)式存在唯一解。

为了用有限元方程解(4.35)式或(4.36)式，先给定Ω一个初始的正规的分划Ω_1；再将Ω_1细分一次得Ω_2；如此继续下去，得网格分划层序列$\Omega_1,\Omega_2,\cdots,\Omega_l$；$\Omega_1$为最粗网格层，$\Omega_l$为最细网格层。

在$\Omega_q(q=1,2,\cdots,l)$网格层上利用三角线性元，可得由线性内插函数构成的试探函数空间V_q^0和与(4.36)式相应的有限元方程组

$$a_q(u_q,v_q)=(f,v_q),\quad\forall v_q\in V_q^0,\quad q=1,2,\cdots,l.\tag{4.37}$$

一、粗网格校正过程

为了在最细网格层Ω_l上解有限元方程

$$a_l(u_l,v_l)=(f,v_l),\quad\forall v_l\in V_l^0,\tag{4.38}$$

设

$$u_l=u_l^j+w_l^j,\quad w_l^j=u_l-u_l^j,\tag{4.39}$$

即(4.38)式的解u_l可看作是近似值u_l^j加上一个校正量w_l^j的结果。

显然，校正量w_l^i满足方程

$$a_l(w_l^i, v_l) = (f, v_l) - a_l(u_l^i, v_l), \quad \forall v_l \in \mathring{V}_l. \tag{4.40}$$

上式右端项也称为残差．据(4.36)式的唯一可解性，(4.40)式右端项可表示为内积形式

$$(r_l^i, v_l^i): = (f, v_l) - a_l(u_l^i, v_l), \quad r_l、v_l \in \mathring{V}_l.$$

称(4.40)式为Ω_l层上的有限元残差方程．

当然对任意给出的近似值u_l^i，若利用(4.40)式解出校正量w_l^i能达到求解(4.38)式的目的，其工作量与用(4.38)式求解并无差别；然而(4.40)式可被看作是校正量w_l^i在$\mathring{V}_l$中的变分问题，

$$J(w_l^i) = a_l(w_l^i, w_l^i) - 2(r_l^i, w_l^i) \Rightarrow \min, \quad w_l^i、r_l^i \in \mathring{V}_l. \tag{4.41}$$

由于$\mathring{V}_{l-1}$是$\mathring{V}_l$的闭子空间，所以w_l^i必定在$\mathring{V}_{l-1}$上取得极小值，即有

$$a_{l-1}(\tilde{\tilde{w}}_l^i, v_{l-1}) = (\tilde{\tilde{r}}_l^i, v_{l-1}), \quad \forall v_{l-1} \in \mathring{V}_{l-1}, \tag{4.42}$$

其中$\tilde{\tilde{w}}_l^i$和$\tilde{\tilde{r}}_l^i$分别表示w_l^i和r_l^i在$\mathring{V}_{l-1}$上的限制．若引进限制算子I_l^{l-1}和线性内插算子I_{l-1}^l，即

$$I_l^{l-1}: V_l \to V_{l-1}, \quad I_{l-1}^l: V_{l-1} \to V_l,$$

则有

$$I_l^{l-1} w_l^i = w_{l-1}^i \in \mathring{V}_{l-1}, \tag{4.43}$$

从而(4.42)式可改写为

$$a_{l-1}(I_l^{l-1} w_l^i, v_{l-1}) = (I_l^{l-1} r_l^i, v_{l-1}),$$

即

$$a_{l-1}(w_{l-1}^i, v_{l-1}) = (r_{l-1}^i, v_{l-1}), \quad \forall v_{l-1} \in \mathring{V}_{l-1}. \tag{4.44}$$

若由(4.44)式解出w_{l-1}^i，则能求解有限元方程组(4.38)，即有

$$u_l = u_l^i + I_{l-1}^l w_{l-1}^i = u_l^i + w_l^i.$$

总而言之，利用解粗网格层Ω_{l-1}上的残差方程(4.44)，可以求得细网格层Ω_l上的(4.38)式的解．这一过程称为有限元粗网格校正过程．应该看到，解(4.44)式的工作量远小于解(4.38)式的

工作量。

为了便于计算和分析，常将(4.37)式写成相应的有限元代数方程组的形式

$$L_q u_q = F_q, \quad q = 1, 2, \cdots, l, \tag{4.45}$$

式中 $L_q = (a_q(\varphi_i^q, \varphi_j^q))_{N_q \times N_q}$， $F_q = ((f, \varphi_j^q))_{N_q \times 1}$，

φ_i^q 和 φ_j^q 是 V_h^q 中的节点基函数，N_q 是 Ω_q 中的节点个数。注意(4.45)式和相应的差分方程组在形式上没有区别，因此人们可以很方便地采用如同本章 §2-§3 所述的多层网格法解有限元方程。这样，多层网格方法可以抽象地超脱有限元方法的物理特征，而只属于对离散微分方程施行该方法的一般性框架，所以多层网格方法有较广的适应性。

5.2 有限元算子"逼近特性"的估计

让我们回忆一下有限元方法涉及的几个关系。二阶椭圆型方程

$$Lu = f \tag{4.46}$$

的弱解方程是

$$a(u, v) = (f, v), \quad \forall v \in H_0^1(\Omega),$$

这里 L 是**自共轭算子**且符合下列关系：

$$a(u, v) = (Lu, v), \quad \forall v \in H_0^1(\Omega), \tag{4.47}$$

$$(Lu, v) = (u, L^* v) = (u, Lv), \tag{4.48}$$

$(\cdot, \cdot)$是通常的L_2内积。

有限元方法是先构造一个有限维连续函数空间序列 $\{V_l\}$ 逼近$H_0^1(\Omega)$，即

$$V_1 \subset V_2 \subset \cdots \subset V_l \subset \cdots \subseteq H_0^1(\Omega),$$

$V_l = \text{span}\{\varphi_i(x) \mid$在$i$的支集上$\varphi_i(x)$局部非零，$i = 1, 2, \cdots, N_l\}$；在有限维连续函数空间$V_l$中得有限元方程：

$$a(u_l^h(x), v_l^h(x)) = (f(x), v_l^h(x)), \quad \forall v_l^h(x) \in V_l; \tag{4.49}$$

解 (4.49) 式得有限元解 $u_l^h(x) \in V_l$ 的离散系数矢量 u_l。通常

(4.49)式可简记为

$$L_l u_l = f_l. \tag{4.50}$$

若记与V_l相应的离散系数矢量的全体为$\mathscr{U}_l$，并引进线性算子 Q_l 和 Q^l，则可建立V_l和$\mathscr{U}_l$之间的联系：

$$Q_l:\ \mathscr{U}_l \to V_l,\ u_l^h(x) = Q_l u_l,$$

$$Q^l:\ V_l \to \mathscr{U}_l,\ u_l = Q^l u_l^h(x).$$

若再定义连续函数内积和离散网格函数内积之关系：

$$\langle Q^l v_l^h(x), w_l\rangle = (v_l^h(x), Q_l w_l), \tag{4.51}$$

其中，$v_l^h(x)\in V_l$，$w_l\in\mathscr{U}_l$，$\langle x_l, y_l\rangle = h_l^2\sum_P x_l^h(P)\cdot\bar{y}_l^h(P)$，$P$ 是网格节点，则在(4.51)式的意义下，Q^l 和Q_l 互为**伴随算子**(参见(4.48)式)：

$$Q^l = Q_l^*,\qquad Q_l = (Q^l)^*. \tag{4.52}$$

还可证明L与L_l、f与f_l有如下对应关系：

$$L_l = Q^l L Q_l,\qquad f_l = Q^l f. \tag{4.53}$$

事实上，利用(4.48)、(4.49)、(4.51)和(4.52) 式，对$\forall v_l^h(x)\in V_l$和$\forall v_l\in\mathscr{U}_l$，一方面有

$$a(u_l^h(x), v_l^h(x)) = (f, v_l^h(x)) = (f, Q_l v_l)$$

$$= \langle Q_l^* f, v_l\rangle = \langle f_l, v_l\rangle = \langle L_l u_l, v_l\rangle,$$

另一方面有

$$a(u_l^h(x), v_l^h(x)) = a(Q_l u_l, Q_l v_l)$$

$$= (L Q_l u_l, Q_l v_l) = \langle Q^l L Q_l u_l, v_l\rangle,$$

故有$L_l = Q^l L Q_l$. 同样易证$f_l = Q^l f$.

分析有限元多层网格方法中线性限制、内插算子I_l^{l-1}、I_{l-1}^l和线性算子Q_l、Q^l之关系. 因

$$Q_{l-1}v_{l-1}\in V_{l-1}\subset V_l,\qquad Q_l(I_{l-1}^l v_{l-1})\in V_l,$$

且 $Q_{l-1}v_{l-1} = Q_l(I_{l-1}^l v_{l-1})$ 表示同一连续函数，故 $Q_{l-1} = Q_l I_{l-1}^l$. 再利用(4.52)式就有$Q^{l-1} = (Q_l I_{l-1}^l)^*$. 总之，如果选择

$$(I_{l-1}^l)^* = I_l^{l-1}\text{或}I_{l-1}^l = (I_l^{l-1})^*, \tag{4.54}$$

则有

$$Q_{l-1}=Q_l I_{l-1}^l, \qquad Q^{l-1}=I_l^{l-1}Q^l. \tag{4.55}$$

为了证明有限元"逼近特性",需注意方程（4.49）的解适合误差估计式：

$$\| u(x)-u_l^h(x)\| \leqslant c h_l^\alpha \| f\|. \tag{4.56}$$

还需注意在(4.51)式意义下，有限维空间中连续函数和离散系数矢量的等价关系：

$$\frac{1}{c}\| v_l\| \leqslant \| Q_l v_l\| \leqslant c\| v_l\|, \qquad \forall v_l \in \mathscr{U}_l.$$

综合前面的结果，下面证明有限元"逼近特性"。

由(4.50)和(4.53)式，有 $u_l^h(x)=Q_l u_l=Q_l L_l^{-1} f_l=Q_l L_l^{-1} Q^l f$，所以由(4.56)式有

$$\| Q_l L_l^{-1} Q^l - L^{-1}\| \leqslant c h_l^\alpha,$$

同样有

$$\| Q_{l-1} L_{l-1}^{-1} Q^{l-1} - L^{-1}\| \leqslant c h_{l-1}^\alpha,$$

利用上面两式和(4.55)式，则有

$$\begin{aligned} & \| (Q_{l-1}L_{l-1}^{-1}Q^{l-1}-L^{-1})-(Q_l L_l^{-1}Q^l-L^{-1})\| \\ =& \| Q_l I_{l-1}^l L_{l-1}^{-1} I_l^{l-1} Q^l - Q_l L_l^{-1} Q^l\| \\ =& \| Q_l (L_l^{-1}-I_{l-1}^l L_{l-1}^{-1} I_l^{l-1}) Q^l\| \leqslant c(h_l^\alpha + h_{l-1}^\alpha) \leqslant c h_l^\alpha. \end{aligned}$$

总上所述，有下述定理。

定理4.6 若线性限制和内插算子互为共轭算子，则有限元"逼近特性"成立：

$$\| L_l^{-1} - I_{l-1}^l L_{l-1}^{-1} I_l^{l-1}\| \leqslant c h_l^\alpha.$$

5.3 有限元方程迭代算子的"光滑特性"分析

设线性有限元方程为

$$L_h u_h = f_h, \tag{4.57}$$

I_h 为 $N\times N$ 阶单位阵，$D_h=d\cdot I_h$ 是对角阵，则解（4.57）式 ω-Jacobi迭代近似解误差 $v_h^k=u_h-u_h^k$ 适合关系：

$$L_h v_h^{k+1} = L_h(I_h - \omega D_h^{-1} L_h) v_h^k = L_h(I_h - \omega D_h^{-1} L_h)^{k+1} v_h^0.$$

显然，不等式

$$\| L_h(I_h - \omega D_h^{-1} L_h)^k \| \leqslant p \| L_h \|, \quad p \in R, \quad 0 < p < 1 \tag{4.58}$$

中的 p 是对(4.57)式施行 k 步 ω-Jacobi迭代的光滑因子，为了分析光滑因子与迭代步数 k 的关系，引进矩阵模：

$$\| Q \| = \sup\{\sqrt{\lambda} \mid \lambda\text{是}Q^*Q\text{的特征值}\}, \tag{4.59a}$$

$$\rho(Q) = \sup\{|\lambda| \mid \lambda\text{是}Q\text{的特征值}\}, \tag{4.59b}$$

和Frobenius模

$$\| Q \|_F = \left(\sum_{i,j=1}^{N} |Q_{ij}|^2 \right)^{1/2},$$

式中 $Q = (Q_{ij})_{N \times N}$，且当 $Q = Q^*$ 时，(4.59a)和(4.59b)式是等价的。

定理4.7 设 L_h 是线性有限元方程的系数矩阵，L_h 对称且 $D_h = d \cdot I_h$，则 ω-Jacobi迭代算子具有“光滑特性”：

$$\| L_h(I_h - \omega D_h^{-1} L_h)^k \| \leqslant \frac{d}{\omega} T(k) \quad (k \geqslant 1). \tag{4.60}$$

k 步迭代的光滑因子 p 满足关系

$$T(k) \leqslant \frac{p\omega}{d} \frac{\| L_h \|_F}{\sqrt{N}} \quad (k \geqslant 1), \tag{4.61}$$

其中 $T(k) = k^k/(1+k)^{k+1}$。

证明 由 $D_h = d \cdot I_h$ 和 L_h 的对称性知 $L_h(I_h - \omega D_h^{-1} L_h)$ 也对称，那么可用 $\| \cdot \| = \rho(\cdot)$ 作算子模，所以有

$$\| L_h(I_h - \omega D_h^{-1} L_h)^k \| = \left\| L_h \left(I_h - \frac{\omega}{d} L_h \right)^k \right\|.$$

令 $0 < \omega \leqslant d / \| L_h \|$，$\lambda$ 是 L_h 的特征值，即有 $0 < \lambda \leqslant d/\omega$，故有

$$\| L_h(I_h - \omega D_h^{-1} L_h)^k \| \leqslant \sup\{\lambda(1 - d^{-1}\omega\lambda)^k \mid 0 < \lambda \leqslant d/\omega\}.$$

令 $x = d^{-1}\omega\lambda$，$x \in (0, 1]$，则上式可改写为

$$\| L_h(I_h - \omega D_h^{-1} L_h)^k \| \leqslant \sup\left\{ \frac{d}{\omega} x(1-x)^k \mid 0 < x \leqslant 1 \right\}.$$

易知当$x=1/(1+k)$时，函数$x(1-x)^k$取得极大值$T(k)=k^k/(1+k)^{k+1}$，所以(4.60)式成立。

另一方面，因L_h的Frobenius模和谱模符合不等式

$$\frac{1}{\sqrt{N}}\|L_h\|_F\leqslant\|L_h\|\leqslant\|L_h\|_F, \tag{4.62}$$

所以由(4.58)式和(4.60)式知，若

$$\|L_h(I_h-\omega D_h^{-1}L_h)^k\|\leqslant\frac{d}{\omega}T(k)\leqslant\frac{p}{\sqrt{N}}\|L_h\|_F\leqslant p\|L_h\|$$

成立，则可从中求得k步迭代的光滑因子p，证得(4.61)式。

最后，利用$0<\lambda\leqslant d/\omega$和(4.68)式，有

$$\frac{1}{\sqrt{N}}\|L_h\|_F\leqslant\max|\lambda|\leqslant\frac{d}{\omega},$$

从中可确定$\omega\leqslant d\sqrt{N}/\|L_h\|_F$。 □

对一致网格剖分，定理4.7的$\|L_h\|_F$可估算出，也可估算出ω的佳值，从而可推算出满足光滑性要求所必须的松弛步k。

例如，对等腰直角三角线性元，第一章中给出了有限元方程的系数值L_h。所以，$\|L_h\|_F=\sqrt{20N}$，$d=4$，$0<\omega\leqslant0.9$，此时与(4.60)式和(4.61)式相应有

$$\|L_h(I_h-\omega D_h^{-1}L_h)^k\|\leqslant\frac{4}{\omega}T(k)\leqslant\frac{4}{\omega}1.118p\omega \quad (k\geqslant1).$$

又如正方形双线性元，$\|L_h\|_F=\sqrt{72N}/3$，$d=8/3$，$0<\omega\leqslant0.95$，此时与(4.60)式和(4.61)式相应有

$$\|L_h(I_h-\omega D_h^{-1}L_h)^k\|\leqslant\frac{8}{\omega}T(k)\leqslant\frac{8}{\omega}1.06p\omega \quad (k\geqslant1).$$

最后，利用等边三角线性元情形说明定理4.7的应用[23]。

对等边三角线性元，$\|L_h\|_F=\sqrt{14N}$，$d=2\sqrt{3}$，$0<\omega\leqslant0.93$；和(4.60)、(4.61)式相应有

$$\|L_h(I_h-\omega D_h^{-1}L_h)^k\|\leqslant\frac{2\sqrt{3}}{\omega}T(k),\quad k\geqslant1, \tag{4.63}$$

$$T(k)\leqslant1.08p\omega,\quad k\geqslant1. \tag{4.64}$$

(4.63)式和(4.64)式说明了下面几个问题。

1　用ω-Jacobi 方法作为光滑迭代，它满足由(4.25)式所定义的“光滑特性”，光滑效果与 h 无关。

2　对于确定的$0<\omega\leqslant 0.93$，利用(4.64)式和$T(k)$可以推算出要满足某种光滑性要求所需要的迭代步数k。

3　由定理证明过程及(4.64)式可知，光滑过程中松弛参数取$\omega\approx 0.93$较理想。

4　利用(4.64)式可作出ω-Jacobi 迭代光滑效果与迭代步之间的关系图，还可作出相邻迭代步压缩误差效果的比值表，见图4.9和表4.5，由此可清楚地看出，ω-Jacobi 法的光滑效果随着迭代步增加而单调变慢，这是与实际计算效果一致的。尽管对高精度的误差要求来说它需要很多迭代步，但前3步的光滑效果较好，

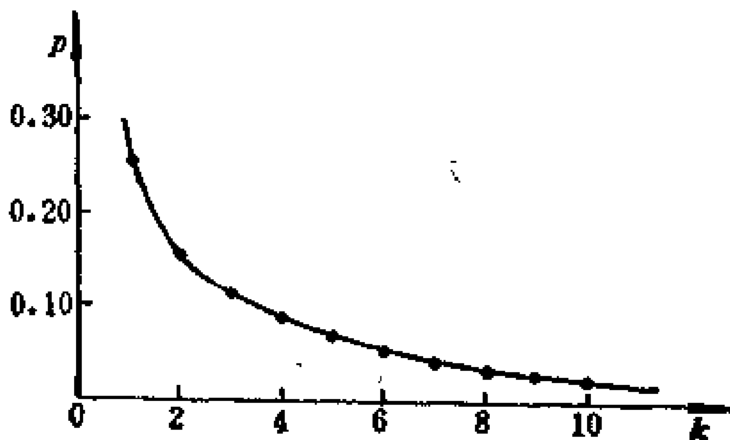


图 4.9

表4.5　Jacobi方法相邻迭代步压缩误迭效果的比法表

k	1	2	3	4	5	6	7	8	9	10
$T(k)/T(k+1)$	1.69	1.41	1.29	1.22	1.18	1.15	1.13	1.12	1.10	1.09

关于有限元多层网格方法中 GS 法的光滑效果，可利用定理4.5和定理4.7中的有关结论得到，即利用关系式

$$\|L_h S_h^k\|\leqslant T(k)\leqslant p\|L_h\|_F/\sqrt{N},$$

其中，当$k=1,2$时，$T(k)=2d/3^k$；

当$k\geqslant 3$时，$T(k)=\frac{1}{2}(k-2)^{\frac{k-2}{2}}/k^{k/2}$。读者可自行确定各种线性元有关$p$和$k$之关系，这样可看出，GS 法的光滑效果也与$h$无关，且在多层网格方法中1～4GS步有较好的效果。

作为本章结束我们指出，文献[14]关于多层网格方法h无关收敛性的证明也是独具特色的，由于篇幅限制，本书未作介绍。另外，多层网格方法也可用于求解非线性离散方程$L_l u_l=f_l$，这里L_l是非线性离散算子。

显然，如果采用某种线性化方法(如Newton 方法)求解非线性离散方程，那么多层网格方法可用于求解线性化方程，这是多层网格方法在求解非线性离散方程方面的间接应用。

多层网格方法也可直接应用于求解非线性离散方程。这里以基本的FAS(full approximation scheme)两层网格方法为例说明之。**FAS两层网格方法**可图解如下：

$$u_h^l\to\xrightarrow[\text{Relax}]{\nu_1}\bar{u}_h^l\to\bar{d}_h^l \qquad \hat{v}_h^l\to\bar{u}_h^l+\hat{v}_h^l\xrightarrow[\text{Relax}]{\nu_2}\to u_h^{l+1}$$

$$\downarrow \hat{I}_h^H\ \downarrow I_h^H \qquad\qquad \uparrow I_H^h$$

$$\bar{u}_H^l\quad \bar{d}_H^l\longrightarrow L_H(\bar{u}_H^l+\hat{v}_H^l)-L_H\bar{u}_H^l=\bar{d}_H^l$$

其中，Relax表示具有光滑作用的非线性松弛方法。与求解线性离散方程的两层网格方法不同，FAS 两层网格方法中，不仅要将$\bar{d}_h^l$限制到粗网格层Ω_H上(可用某个线性限制算子I_h^H)，而且还要将$\bar{u}_h^l$限制到粗网格层上(用某个线性限制算子$\hat{I}_h^H$，它可以不同或相同于I_h^H)；这相当于在Ω_h上定义的残差方程是

$$L_h(\bar{u}_h^l+v_h^l)-L_h\bar{u}_h^l=\bar{d}_h^l,$$

在Ω_H上的类似残差方程是

$$L_H(\bar{u}_H^l+\hat{v}_H^l)-L_H\bar{u}_H^l=\bar{d}_H^l,$$

或等价于

$$L_H w_H^l=\bar{d}_H^l+L_H\bar{u}_H^l,\quad \hat{v}_H^l=w_H^l-\bar{u}_H^l;$$

也就是说，FAS两层网格方法不是在 Ω_H 上直接求解校正量，而是先求w_H^l再求校正量 $\hat{v}_H^l$；这种做法不需要线性化过程就可方便地数值求解非线性离散方程；如果L_h是线性离散算子则FAS 两层网格方法与一般两层网格方法(见第四章§2.1)是一致的。FAS两层网格方法很容易推广为FAS多层网格方法。

多层网格方法的应用是广泛的。虽然本章中只是就模型问题的第一边值情形为例，但多层网格方法也同样有效地用于求解其它混合边值问题，特征值数值求解问题，积分方程数值求解问题。更进一步地，在其它种类的数值求解中，不一定局限于微分方程的数值求解，只要能建立一个“粗”、“细”不同的离散化问题的序列去有效地逼近原来的连续性问题，人们就可用多层网格方法的计算原则去达到提高计算效率的目的。

第五章　逐层子空间迭代法

高效率高精度计算微分方程数值解是目前正在广泛深入研究的课题。第四章中介绍的多层网格方法目前已得到了进一步的发展和应用。多层网格方法是求解离散微分方程的一种高效率算法，但由该方法获得的数值解精度仍同于使用经典迭代法的求解精度，换句话说，多层网格方法并不能改变由离散方程确定的数值解精度阶。陈传淼和朱起定等人研究了有限元解的超收敛性，为提高有限元导数值精度作出了很有理论和实用价值的贡献。林群、吕涛和陈传淼等人研究了有限元解的渐近展开式，不仅在理论上深入研究了有限元解的性质，而且提出了实用的外推公式，从而为提高有限元位移值的精度作出了极有价值的贡献，使线性元发挥了高次元的作用。以上这三方面的研究都是从不同的侧面单方面进行的。本章将综合他们的研究成果，把外推、超收敛性、分裂算法和多层网格方法有机地结合起来，提出一种高效率、高精度求解离散微分方程(有限元方程和差分方程)的算法[90]，并称之为“逐层子空间迭代方法”(简记为I.S.S.方法)。I.S.S.方法不仅保持了多层网格方法的高效率，保持了外推、超收敛性技术的高精度效果，而且能加速数值求解过程，大大地节省计算量。I.S.S.方法能达到人们所追求的几个指标：(1)存贮量达最佳水平；(2)计算量达$O(N_l)$，N_l为细网展节点的个数；(3)敛速与h无关；(4)数值解精度高；(5)适应性广，既可用于求解差分方程和有限元方程，也可用于数值求解线性或非线性的椭圆型边值问题和抛物型初边值问题等；(6)有利于计算机实现网格剖分、计算、输人和输出的全自动化过程，特别地对于一般三角形单元和任意四边形单元剖分情形，能以极少的信息输入量获得所需要

的理想的效果；(7)便于事后误差估计。

本章§6中还给出了一些用 I.S.S.方法求解模型问题的数值结果，实例计算结果表明，使用 I.S.S.算法比使用多层网格方法(V循环或W循环等)及带有最佳松弛因子的 SOR 方法等的效果要好，不仅其计算工作量可大量节省，而且数值解精度也有明显的提高。

§1 逐层子空间迭代法的计算步骤

设 $\Omega_1, \Omega_2, \cdots, \Omega_l$ 是给定的网格层序列，Ω_1 和 Ω_l 分别是最粗和最细网格层，$\Omega_i(i=1,2,\cdots,l)$网格层的网格参数为 h_i，这些网格层之间满足下列关系：

$$\begin{cases} \Omega_i \subset \Omega_{i+1}, \quad \mathrm{Dim}(\Omega_i) \ll \mathrm{Dim}(\Omega_{i+1}), \quad i=1,2,\cdots,l-1, \\ \Omega_{i+1} = \Omega_i \cup (\Omega_{i+1}/\Omega_i), \\ \Omega_l = \Omega_1 \cup (\Omega_2/\Omega_1) \cup \cdots \cup (\Omega_l/\Omega_{l-1}). \end{cases} \tag{5.1}$$

为了数值求解线性偏微分方程的边值问题

$$\begin{cases} Lu = f, & (x,y) \in \Omega, \\ u = g, & (x,y) \in \partial\Omega, \end{cases}$$

选用某种离散化方法(有限元方法或差分方法)就得到定义在Ω_i网格层上的离散微分方程：

$$L_i u_i = f_i, \quad i=1,2,\cdots,l, \tag{5.2}$$

式中，L_i 是定义在 Ω_i 网格层上的离散算子，u_i 和 f_i 是定义在 Ω_i 网格层上的网格函数的系数矢量。为了特殊需要，它们有时也记为 $u_i(\Omega_i)$ 和 $f_i(\Omega_i)$，显然，

$$u_i(\Omega_i) = u_i(\Omega_{i-1}) \cup u_i(\Omega_i/\Omega_{i-1}), \quad i=2,3,\cdots,l.$$

为了在最细网格层 Ω_l 上高效率高精度地获得离散微分方程(5.2)的数值解$u_l(\Omega_l)$，**逐层子空间迭代法**(I.S.S.)的计算步骤如下：

第一步，解 $L_i u_i = f_i$，$i=1,2$，得 $u_1(\Omega_1)$ 和 $u_2(\Omega_2)$；

第二步，利用 $u_1(\Omega_1)$ 和 $u_2(\Omega_2)$施行外推技巧（参见§3），得$\tilde{u}_2(\Omega_2)$，并令 $\tilde{u}_3(\Omega_2)=\tilde{u}_2(\Omega_2)$；

第三步，利用$\tilde{u}_3(\Omega_2)$内插得$\hat{u}_3(\Omega_3/\Omega_2)$；

第四步，利用$\hat{u}_3(\Omega_3/\Omega_2)$作初值，在 Ω_3/Ω_2 上作子空间迭代（参见§4）得$\bar{u}_3(\Omega_3/\Omega_2)$；

第五步，以$\tilde{u}_3(\Omega_2)\cup\bar{u}_3(\Omega_3/\Omega_2)$为初值，在 Ω_3 网格层上施

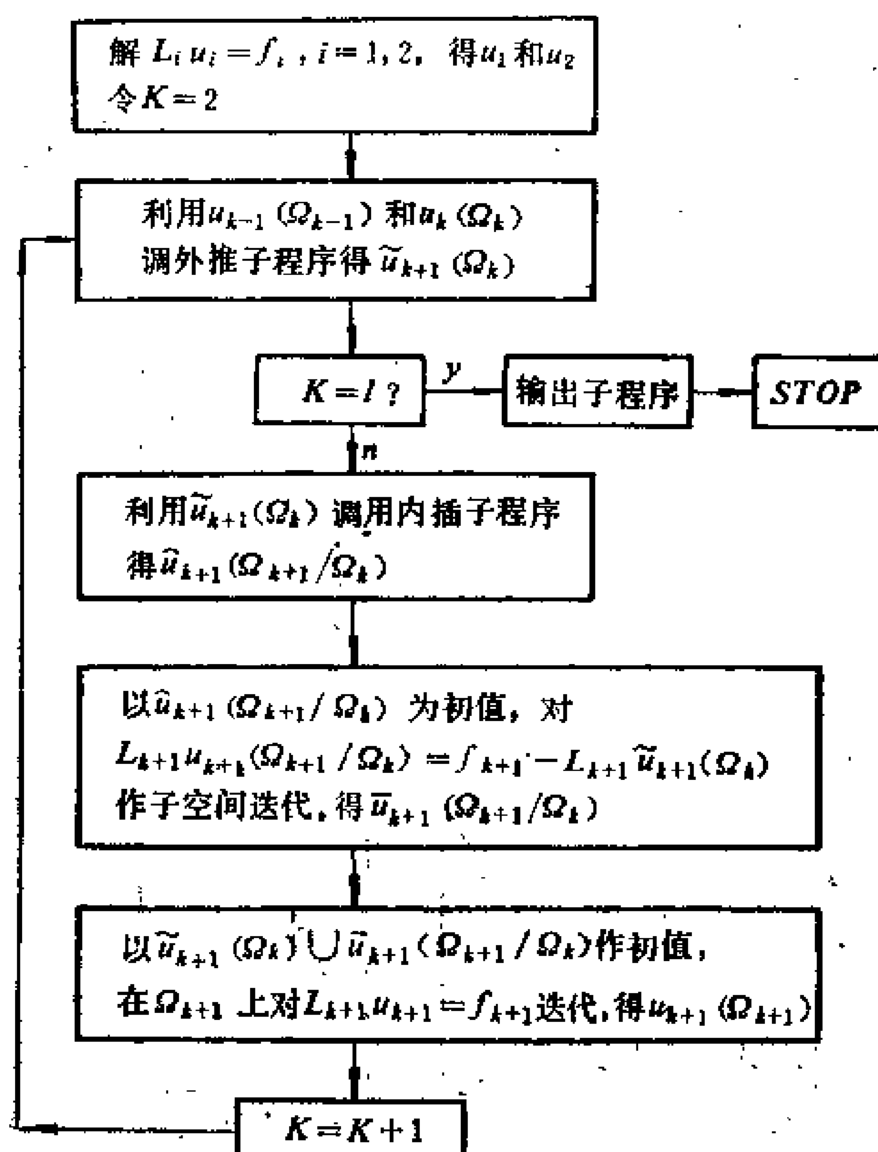


图 5.1

行 ν 步松弛迭代，得 $u_3(\Omega_3)$，即，

$$u_3(\Omega_3) = \text{Relax}\nu(\hat{u}_3(\Omega_2) \cup \bar{u}_3(\Omega_3/\Omega_2),\ L_3, f_3, \omega);$$

第六步，在 $\Omega_k(k=3,4,\cdots,l)$ 网格层上，重复如同第二步到第五步的计算过程，直至求出 $u_l(\Omega_l)$；

第七步，最后在 Ω_l(利用 Ω_{l-1})网格层上再使用一次外推和超收敛性技巧，就可获得高精度的离散位移值和导数值。

I.S.S.算法简单易行，图 5.1 表示了该算法的简化计算流程，这里需要两组存贮单元。

若记"→"、"↑"、"□"和"▨"分别表示外推过程、内插过程、经外推技巧后的获值集合和要求解的子集合，则图5.2表示了I.S.S.算法的实际计算过程的图解。

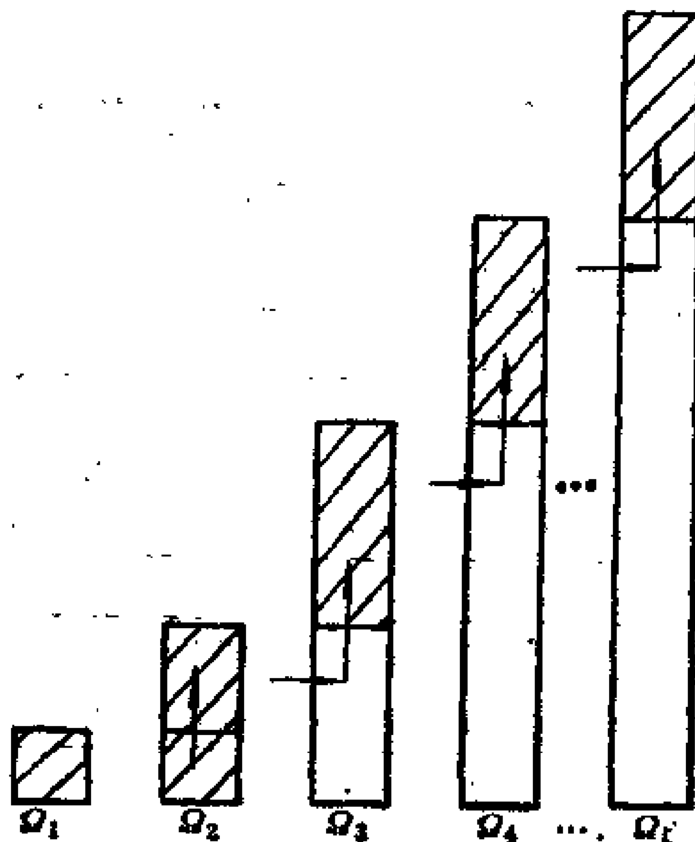


图 5.2

§2　网格序列的构造

先考虑定解区域是矩形的简单情形。不论是采用直角三角形剖分，还是矩形剖分，离散节点都可统一采用坐标式编号，节点记为(i,j).设最细网格层是这样得到的：将一条矩形边均分2^m分，另一条均分2^n分；则可构造出网格层序列$\{\Omega_i\}(i=1,2,\cdots,l)$，$h_l$是网格参数，$h_l=(h_{l1},h_{l2})$，$h_l=2h_{l+1}$。在连同边界点一起考虑时，各网格层节点编号的规律如下：

$$\begin{cases}\Omega_l=\left\{(i,j)\middle|\begin{matrix}i=1,2^m+1,\ \text{step}2^0\\ j=1,2^n+1,\ \text{step}2^0\end{matrix}\right\},\\ \Omega_{l-1}=\left\{(i,j)\middle|\begin{matrix}i=1,2^m+1,\ \text{step}2^1\\ j=1,2^n+1,\ \text{step}2^1\end{matrix}\right\},\\ \cdots\\ \Omega_2=\left\{(i,j)\middle|\begin{matrix}i=1,2^m+1,\ \text{step}2^{l-2}\\ j=1,2^n+1,\ \text{step}2^{l-2}\end{matrix}\right\},\\ \Omega_1=\left\{(i,j)\middle|\begin{matrix}i=1,2^m+1,\ \text{step}2^{l-1}\\ j=1,2^n+1,\ \text{step}2^{l-1}\end{matrix}\right\},\end{cases}\tag{5.3}$$

其中，$l=\min(m,n)$。此时，相邻两网格层的节点分布情况如图

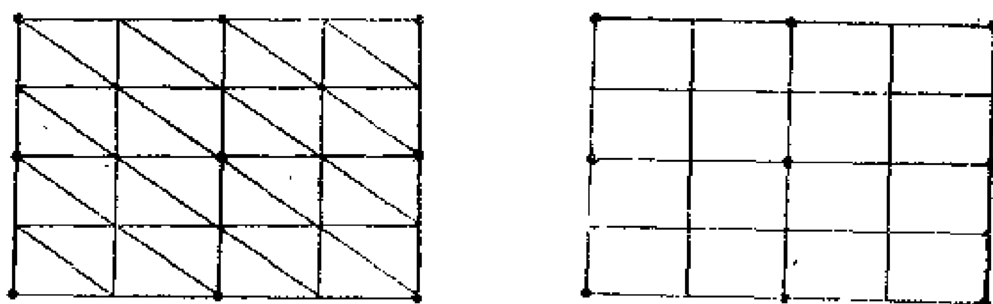
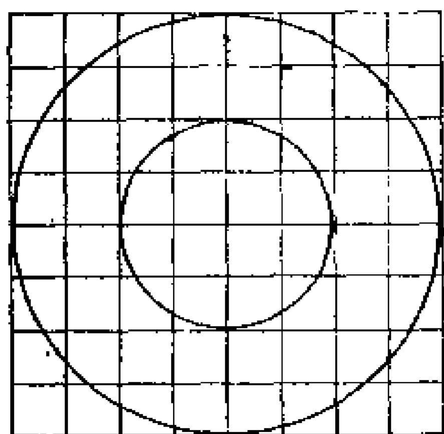


图5.3　•属于Ω_l，全部节点属于Ω_{l+1}

5.3所示。显然，由(5.3)式定义的网格层序列适合关系式(5.1)。值得指出的是，在网格节点采用坐标式编号的情况下，不论采用差分方法或有限元方法，离散微分方程(5.2)的系数阵L_l的元素

都非常规则，因此在实际计算中，使用局部离散格式特别方便，L_l的元素不必存贮。

再考虑曲边定解区域的矩形网格分划情形，此时仍可采用坐标式编号，为此，人们可以设想一个如（5.3）式所定义的网格层序列，覆盖住曲边区域Ω，参见图 5.4．虽然此时节点编号有某些规则可循，但又产生了几个新的问题．第一，要区分哪些网格节点在界外或在界内，因为界外网格节点并不参与运算．第二，要确定网格线与边界的交点．第三，对界内网格节点，要区分哪些是“正则内点”，哪些是“非正则内点”；对于“正则内点”，可采用非常规则的局部离散格式；对于“非正则内点”，其局部离散格式是不规则的，所以必须存贮那些在运算过程中涉及到的有关信息．如对图 5.5 中的“非正则内点”Q采用不等距离散：


图 5.4

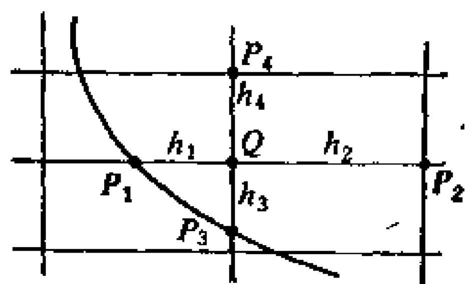


图 5.5

$$\left\{\begin{aligned}
&\frac{\partial u}{\partial x}(Q)\approx\frac{h_2}{h_1+h_2}\left[\frac{u(Q)-u(P_1)}{h_1}\right]+\frac{h_1}{h_1+h_2}\left[\frac{u(P_2)-u(Q)}{h_2}\right],\\
&\frac{\partial u}{\partial y}(Q)\approx\frac{h_4}{h_3+h_4}\left[\frac{u(Q)-u(P_3)}{h_3}\right]+\frac{h_3}{h_3+h_4}\left[\frac{u(P_4)-u(Q)}{h_4}\right],\\
&\frac{\partial^2 u}{\partial x^2}(Q)\approx\frac{2}{h_1+h_2}\left[\frac{u(P_2)-u(Q)}{h_2}-\frac{u(Q)-u(P_1)}{h_1}\right],\\
&\frac{\partial^2 u}{\partial y^2}(Q)\approx\frac{2}{h_3+h_4}\left[\frac{u(P_4)-u(Q)}{h_4}-\frac{u(Q)-u(P_3)}{h_3}\right],
\end{aligned}\right.\tag{5.4}$$

则必须存贮 h_1、h_2、h_3 和 h_4。

显然，曲边形区域的这种离散化方式也同样形成了一个符合(5.1)式要求的网格层序列。这样的区域离散化方式还具有下述三个优点：其一，由于采用了坐标式编号，"正则内点"处采用规则的局部离散格式，所以计算离散方程系数阵元素的工作量和存贮量都很少，从而避免了在一般三角形分划的有限元方法情形需大量计算和存贮的缺点。其二，有利于采用不同的离散化格式，例如"非正则内点"处可采用不等距 5 点差分格式，也可采用三角线性元和三角二次元格式；在"正则内点"处可采用差分格式，也可采用有限元格式。其三，可顺利地扩展到三维情形。

对于曲边定解区域的一般三角形分划情形，在按照第二篇第三章§3中介绍的机器自动剖分办法执行时，机器可自动生成符合(5.1)式要求的网格层序列。

§3 外推和内插公式

文献[17]和[26]研究了在网格步长分半减小时，数值解精度不断提高的规律，提出一些外推公式，并在实践中取得了一定的效果。近年来，有限元近似解渐近展开式的研究[17]~[26],[30]不仅在理论上深入讨论了离散解与准确解之间的关系，有重要的意义；而且在实际应用方面提供了使用外推技巧提高数值解精度的广阔前景。

现已证明，在 Ω_m 网格层上(网格参数为 h_m)不论采用 5 点或 9 点差分格式，还是采用三角线性元格式或矩形双线元格式等；不论是数值求解线性或拟线性椭圆型微分方程，还是积分方程；只要微分方程的解 u 适当光滑，数值解 u_m 和内插函数 u_I 就具有如下关系：

$$u_m(Z)=u_I(Z)+\phi(Z)h_m^{\alpha}+O(h_m^{\alpha+1}),\quad \forall Z\in\Omega_m, \tag{5.5}$$

其中 $\phi(Z)$ 是与 h_m 无关的函数。

仿照第二篇第四章§1中的有关推导，当 $h_m=2h_{m+1}$ 时，就可得到下面一组外推公式：

采用三角形分划，P是单元节点，M是单元边中点，M_i 是M的同边邻节点，

$$\begin{cases} u_{m+1}^{**}(P)=u_{m+1}(P)+\dfrac{1}{2^\alpha-1}(u_{m+1}-u_m)(P) \\ \qquad\qquad =u(P)+O(h_{m+1}^{\alpha+1}), \\ u_{m+1}^{**}(M)=u_{m+1}(M)+\dfrac{1}{2^{\alpha+1}-2}\sum\limits_{i=1}^{2}(u_{m+1}-u_m)(M_i) \\ \qquad\qquad =u(M)+O(h_{m+1}^{\alpha+1}); \end{cases} \tag{5.6}$$

采用矩形分划，P 是单元节点，M是单元边中点，M_i 是M的同边邻节点，Q 是单元中心，Q_i 是Q的同单元顶节点，

$$\begin{cases} u_{m+1}^{**}(P)=u_{m+1}(P)+\dfrac{1}{2^\alpha-1}(u_{m+1}-u_m)(P) \\ \qquad\qquad =u(P)+O(h_{m+1}^{\alpha+1}), \\ u_{m+1}^{**}(M)=u_{m+1}(M)+\dfrac{1}{2^{\alpha+1}-2}\sum\limits_{i=1}^{2}(u_{m+1}-u_m)(M_i) \\ \qquad\qquad =u(M)+O(h_{m+1}^{\alpha+1}), \\ \qquad\qquad =u_{m+1}^{**}(Q)=u_{m+1}(Q)+\dfrac{1}{2^{\alpha+2}-4}\sum\limits_{i=1}^{4}(u_{m+1}-u_m)(Q_i) \\ \qquad\qquad =u(Q)+O(h_{m+1}^{\alpha+1}). \end{cases} \tag{5.7}$$

若记 $\varepsilon_{m+1}(P)=u_{m+1}(P)-u(P)$，$\delta_{m+1}(P)$是一高阶小量，$P$是单元节点，则由(5.5)式得

$$\varepsilon_{m+1}(P)=\left(\frac{1}{2}\right)^\alpha\varepsilon_m(P)+\delta_{m+1}(P), \tag{5.8}$$

于(5.8)式两边同时加上$\left(\frac{1}{2}\right)^\alpha u_{m+1}(P)$，就有

$$\left[1+\left(\frac{1}{2}\right)^\alpha\right]u_{m+1}(P)-\left(\frac{1}{2}\right)^\alpha u_m(P)-u(P)$$

$$=\left(\frac{1}{2}\right)^\alpha(u_{m+1}(P)-u(P))+\delta_{m+1}(P),$$

这就是说，若令

$$u_{m+1}^*(P)=\left[1+\left(\frac{1}{2}\right)^{\alpha}\right]u_{m+1}(P)-\left(\frac{1}{2}\right)^{\alpha}u_m(P)$$

$$=u_{m+1}(P)+\left(\frac{1}{2}\right)^{\alpha}(u_{m+1}-u_m)(P),$$

则有

$$u_{m+1}^*(P)-u(P)=\left(\frac{1}{2}\right)^{\alpha}\varepsilon_{m+1}(P)+\delta_{m+1}(P). \tag{5.9}$$

另一方面，由(5.8)式有

$$u_{m+2}(P)-u(P)=\left(\frac{1}{2}\right)^{\alpha}\varepsilon_{m+1}(P)+\delta_{m+2}(P),$$

将其与(5.9)式相减，得

$$u_{m+2}(P)-u_{m+1}^*(P)=\delta_{m+2}(P)-\delta_{m+1}(P),$$

再与

$$u_{m+2}(P)-u_{m+1}(P)=\left[\left(\frac{1}{2}\right)^{\alpha}-1\right]\varepsilon_{m+1}(P)+\delta_{m+2}(P)$$

比较，即知 $u_{m+1}^*(P)$ 也比 $u_{m+1}(P)$ 更接近 $u_{m+2}(P)$，或即

$$|u_{m+1}^*(P)-u_{m+2}(P)|<|u_{m+1}(P)-u_{m+2}(P)|. \tag{5.10}$$

因此，$u_{m+1}^*(P)$ 可作为另一种外推值。仍然仿照第一篇第四章§1中的有关推导，就可得到与(5.6)式和(5.7)式相应的另一组外推公式：

采用三角形分划，P 是单元节点，M 是单元边中点，M_i 是 M 的同边邻节点，

$$\begin{cases} u_{m+1}^*(P)=u_{m+1}(P)+\left(\frac{1}{2}\right)^{\alpha}(u_{m+1}-u_m)(P) \\ \qquad =u(P)+\left(\frac{1}{2}\right)^{\alpha}\varepsilon_{m+1}(P)+\delta_{m+1}(P), \\ u_{m+1}^*(M)=u_{m+1}(M)+\left(\frac{1}{2}\right)^{\alpha+1}\sum_{i=1}^{2}(u_{m+1}-u_m)(M_i) \\ \qquad =u(M)+\left(\frac{1}{2}\right)^{\alpha}\varepsilon_{m+1}(M)+\delta_{m+1}(M); \end{cases} \tag{5.11}$$

采用矩形分划，P 是单元节点，M 是单元边中点，M_i 是 M

的同边邻节点，Q是单元中心，Q_i是Q的同单元顶节点，

$$
\begin{cases}
u_{m+1}^{*}(P)=u_{m+1}(P)+\left(\frac{1}{2}\right)^{\alpha}(u_{m+1}-u_m)(P) \\
\qquad =u(P)+\left(\frac{1}{2}\right)^{\alpha}\varepsilon_{m+1}(P)+\delta_{m+1}(P), \\
u_{m+1}^{*}(M)=u_{m+1}(M)+\left(\frac{1}{2}\right)^{\alpha+1}\sum_{i=1}^{2}(u_{m+1}-u_m)(M_i) \\
\qquad =u(M)+\left(\frac{1}{2}\right)^{\alpha}\varepsilon_{m+1}(M)+\delta_{m+1}(M), \\
u_{m+1}^{*}(Q)=u_{m+1}(Q)+\left(\frac{1}{2}\right)^{\alpha+2}\sum_{i=1}^{4}(u_{m+1}-u_m)(Q_i) \\
\qquad =u(Q)+\left(\frac{1}{2}\right)^{\alpha}\varepsilon_{m+1}(Q)+\delta_{m+1}(Q).
\end{cases}
\tag{5.12}
$$

u_{m+1}^{**}、u_{m+1}^{*}和u_{m+1}之间的关系是

$$
|u_{m+1}^{**}(P)-u(P)|<|u_{m+1}^{*}(P)-u(P)|<|u_{m+1}(P)-u(P)|.
\tag{5.13}
$$

事实上，由(5.6)式有

$$
u_{m+1}^{**}(P)=\frac{1}{1-\left(\frac{1}{2}\right)^{\alpha}}\left[u_{m+1}(P)-\left(\frac{1}{2}\right)^{\alpha}u_m(P)\right]
$$

$$
=\frac{1}{1-\left(\frac{1}{2}\right)^{\alpha}}\left[\varepsilon_{m+1}(P)-\left(\frac{1}{2}\right)^{\alpha}\varepsilon_m(P)+\left(1-\frac{1}{2^{\alpha}}\right)u(P)\right],
$$

再由(5.8)式，就有

$$
u_{m+1}^{**}(P)-u(P)=\delta_{m+1}(P)\Big/\left[1-\left(\frac{1}{2}\right)^{\alpha}\right],
$$

将其与(5.9)式比较，即知u_{m+1}^{**}在逐点意义下比u_{m+1}^{*}具有更高的逼近精度。最后还考虑到(5.10)式，即知(5.13)式是正确的。

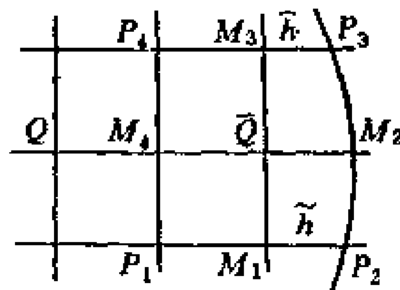


图 5.6

下面介绍不等距外推公式。设定解区域为曲边形，如图5.6所

示；P_4、P_1、M_4 和 Q 诸点是正则内点，M_3、$\bar{Q}$ 和 M_1 是非正则内点；P_1、P_2、P_3 和 P_4 属于 Ω_h，M_1、M_2、M_3 和 $\bar{Q}$ 属于 $\Omega_{h/2}$。为了方便起见，还设 $h_1 = h_2 = h$；设采用具有 α 阶精度的局部离散格式，即渐近展式（5.5）成立；则不难推导出下面不等距外推公式：

$$
\begin{cases}
(5.7)\text{式}, \qquad\qquad\qquad \text{正则内点处}, \\
u_{m+1}^{**}(M_1) = u_{m+1}(M_1) + \dfrac{1}{2^\alpha - 1}[u_{m+1}(P_1) - u_m(P_1)]\dfrac{\tilde{h}}{h+\tilde{h}} \\
\qquad + \dfrac{1}{2^\alpha - 1}[u_{m+1}(P_2) - u_m(P_2)]\dfrac{h}{h+\tilde{h}} \\
\qquad = u(M_1) + O(h^{\alpha+1}), \\
u_{m+1}^{**}(M_3) = u_{m+1}(M_3) + \dfrac{1}{2^\alpha - 1}[u_{m+1}(P_4) - u_m(P_4)]\dfrac{\hat{h}}{h+\hat{h}} \\
\qquad + \dfrac{1}{2^\alpha - 1}[u_{m+1}(P_3) - u_m(P_3)]\dfrac{h}{h+\hat{h}} \qquad\qquad (5.14) \\
\qquad = u(M_3) + O(h^{\alpha+1}), \\
u_{m+1}^{**}(\bar{Q}) = u_{m+1}(\bar{Q}) + \dfrac{1}{2^\alpha - 2}\displaystyle\sum_{i=1}^{4}\varepsilon_i[u_{m+1}(P_i) - u_m(P_i)] \\
\qquad = u(\bar{Q}) + O(h^{\alpha+1}), \\
\varepsilon_1 = \dfrac{\sqrt{h^2+\hat{h}^2}}{\sqrt{2}\,h + \sqrt{h^2+\hat{h}^2}}, \qquad \varepsilon_3 = \dfrac{\sqrt{2}\,h}{\sqrt{2}\,h + \sqrt{h^2+\hat{h}^2}}, \\
\varepsilon_2 = \dfrac{\sqrt{2}\,h}{\sqrt{2}\,h + \sqrt{h^2+\tilde{h}^2}}, \qquad \varepsilon_4 = \dfrac{\sqrt{h^2+\tilde{h}^2}}{\sqrt{2}\,h + \sqrt{h^2+\tilde{h}^2}}.
\end{cases}
$$

本节介绍的两种外推公式都可用于I.S.S.算法中。经I.S.S.算法的外推步骤，就有

$$|\tilde{u}_m(\Omega_m) - u(\Omega_m)|_\infty < |u_m(\Omega_m) - u(\Omega_m)|_\infty,$$

从而提高了 Ω_m 网格层上数值解的精度。外推值 $\tilde{u}_m(\Omega_m)$ 精度较高，可作为 $u_{m+1}(\Omega_{m+1})$ 在 Ω_m 节点处的极好的近似值 $\tilde{u}_{m+1}(\Omega_m)$；

再利用 $\tilde{u}_{m+1}(\Omega_m)$内插得$\hat{u}_{m+1}(\Omega_{m+1}/\Omega_m)$，并经子空间迭代得$\tilde{u}_{m+1}(\Omega_{m+1}/\Omega_m)$（子空间迭代收敛快，工作量小，参见§5），这样 $\tilde{u}_{m+1}(\Omega_m)\cup\tilde{u}_{m+1}(\Omega_{m+1}/\Omega_m)$已基本上接近 $u_{m+1}(\Omega_{m+1})$的精度。因此，外推和子空间迭代大大加速了数值求解进程；当然，网格逐步细化，逐层数值求解并予以嵌套利用也是加速数值求解进程的因素。若只就外推技术而言，它在 I.S.S. 算法中不仅加速了数值求解进程，而且能最终输出高精度数值解，所以外推技术在逐层子空间迭代中起着重要的作用。

外推过程是容易实现的，只要对 Ω_{m+1} 中的不同类型的节点循环并按公式(5.6)和(5.7)式（或公式(5.11)和(5.12)式）分别处理即可。

内插过程是很容易实现的，这在第三章§2中已讨论过了。这里再介绍一种基于 RB 分划下局部离散算子的启发而获得的内插方法。以数值求解 Poisson 方程为例，利用 $\tilde{u}_{m+1}(\Omega_m)$ 内插得 $\hat{u}_{m+1}(\Omega_{m+1}/\Omega_m)$的具体办法是

（1）对于 Ω_m 网格单元的中心点 P，采用计算公式：

$$\frac{1}{2h_{m+1}^2}\begin{bmatrix}-1 & & -1\\ & 4 & \\ -1 & & -1\end{bmatrix}\hat{u}_{m+1}(P)=f(P).$$

这个内插过程是显式计算的，它是在离散方程的启发下得到的。这个离散格式相当于定义在步长为$\sqrt{2}h_{m+1}$ 的 RB 分划网格上的 5 点离散格式。

（2）对于 Ω_m 网格单元的边中点 P，采用计算公式

$$\frac{1}{h_{m+1}^2}\begin{Bmatrix} & -1 & \\ -1 & 4 & -1\\ & -1 & \end{Bmatrix}\hat{u}_{m+1}(P)=f(P).$$

在步骤(1)完成后，上式的计算也是显式的，这个内插过程是直接利用定义在 Ω_{m+1} 网格上的 5 点离散格式得到的。

这种内插办法在数值求解其它的偏微分方程的过程中也是适

用的，只不过要把离散格式作些相应的修改就行了。

§4 子空间迭代与事后误差估计

现假设利用 u_{m-1} 和 u_m 外推得高精度离散值 $\tilde{u}_{m+1}(\Omega_m)$，因为

$$|\tilde{u}_{m+1}(\Omega_m)-u(\Omega_m)|_\infty<|u_{m+1}(\Omega_m)-u(\Omega_m)|_\infty,$$

所以有必要充分利用 $\tilde{u}_{m+1}(\Omega_m)$ 这个高精度值。正如§1中关于I.S.S.算法步骤所述，固定 $\tilde{u}_{m+1}(\Omega_m)$，即令 $\tilde{u}_{m+1}(\Omega_m)=\tilde{u}_m(\Omega_m)$，再内插 $\tilde{u}_{m+1}(\Omega_m)$ 得 $\hat{u}_{m+1}(\Omega_{m+1}/\Omega_m)$，并以此为初值在子空间 Ω_{m+1}/Ω_m 上用松弛法改进 $\hat{u}_{m+1}(\Omega_{m+1}/\Omega_m)$ 的精度，这样经少量迭代步即可获得精度与 $\tilde{u}_{m+1}(\Omega_m)$ 相当的 $\tilde{u}_{m+1}(\Omega_{m+1}/\Omega_m)$。这里，我们把固定 $\Omega_m\cap\Omega_{m+1}$ 上的值，只在子集 Ω_{m+1}/Ω_m 上进行的迭代称之为**子空间迭代**；相应地，在 Ω_{m+1} 整个网格层上进行的迭代称之为**全空间光滑**。

不难理解，如果不作子空间迭代，而只是利用 $\tilde{u}_{m+1}(\Omega_m)\cup\hat{u}_{m+1}(\Omega_{m+1}/\Omega_m)$ 作初值在 Ω_{m+1} 上作全空间光滑，则会花费较多的工作量。这是因为 $\tilde{u}_{m+1}(\Omega_m)$ 的精度较高，$\hat{u}_{m+1}(\Omega_{m+1}/\Omega_m)$ 的精度较差，所以在全空间光滑初期，具有局部加权平均特性的迭代过程使得 $\tilde{u}_{m+1}(\Omega_m)$ 的精度降低，这当然会导致计算量的浪费。而子空间迭代则固定 $\tilde{u}_{m+1}(\Omega_m)$，并强迫 $\hat{u}_{m+1}(\Omega_{m+1}/\Omega_m)$ 迅速提高精度，这样只要花费较少的计算量就可达到 $\tilde{u}_{m+1}(\Omega_m)$ 和 $\tilde{u}_{m+1}(\Omega_{m+1}/\Omega_m)$ 精度匹配的目的。正是由于此，子空间迭代能加速整个数值求解的进程。

子空间迭代是一种特殊的分裂算法。假设 Ω_m 和 Ω_{m+1} 网格层的网格节点总数分别为 N_m 和 N_{m+1}，对应 Ω_{m+1} 网格层上的离散方程为

$$L_{m+1}u_{m+1}=f_{m+1}, \tag{5.15}$$

其中 L_{m+1} 是 $N_{m+1}\times N_{m+1}$ 阶矩阵。若再假定 Ω_m 中的节点先编

号，Ω_{m+1}/Ω_m 中的节点后编号，则对 $\Omega_m\cap\Omega_{m+1}$ 中的任一节点，其全部邻点都属于 Ω_{m+1}/Ω_m 中，故 L_{m+1} 可写成块状结构形式，且有

$$L_{m+1}u_{m+1}(\Omega_{m+1})=$$

$$\begin{bmatrix} A & C^T \\ C & B \end{bmatrix}\begin{bmatrix} u_{m+1}(\Omega_m) \\ u_{m+1}(\Omega_{m+1}/\Omega_m) \end{bmatrix}=\begin{bmatrix} f_{m+1}(\Omega_m) \\ f_{m+1}(\Omega_{m+1}/\Omega_m) \end{bmatrix}, \qquad (5.16)$$

其中，A 是 $N_m\times N_m$ 阶矩阵，B 是 $(N_{m+1}-N_m)\times(N_{m+1}-N_m)$ 阶矩阵，C 是 $(N_{m+1}-N_m)\times N_m$ 阶矩阵。因为 $\tilde{u}_{m+1}(\Omega_m)$ 作强制边值处理，所以可将方程组(5.16)分裂为子空间中的数值求解问题：

$$\begin{cases} u_{m+1}(\Omega_m)=\tilde{u}_{m+1}(\Omega_m), \\ Bu_{m+1}(\Omega_{m+1}/\Omega_m)=f_{m+1}(\Omega_{m+1}/\Omega_m)-C\tilde{u}_{m+1}(\Omega_m). \end{cases} \qquad (5.17)$$

这是一个特殊的分裂算法。因此，在 Ω_{m+1} 上求解(5.16)式就转化为一个低维(维数是 $N_{m+1}-N_m$)方程组(5.17)的求解问题，当然这对节省工作量是有益的。另外，(5.17)式是唯一可解的，这是因为(5.17)式中的矩阵 B 是(5.16)式中的 L_{m+1} 经若干强制性限制后的结果，所以矩阵 B 的正定性要比 L_{m+1} 的好得多，在有限元方程情形关于 B 矩阵正定性的分析可参见[30]。

如按(5.3)式坐标编号规则，虽然此时方程(5.15)的系数阵不再具有块状结构，但仍可以写成(5.17)的形式，只不过其中的 B 和 C 矩阵都是大型稀疏矩阵。因为子空间迭代仍是一个迭代过程，它的有效运算只涉及系数矩阵的非零元素部分，所以在子空间(Ω_{m+1}/Ω_m)上迭代求解(5.15)式仍是一个低维方程组的求解问题。更一般地，可写出(5.15)式中某一个方程的 SOR 迭代计算公式(参见(2.16)式)：

$$u_i^{k+1}=(1-\omega)u_i^k+\frac{\omega}{a_{ii}}\left(f_i-\sum_{j\in V^+}a_{ij}u_j^{k+1}-\sum_{j\in V^-}a_{ij}u_j^k\right),$$

在子空间迭代过程中，若判得 $j\in\Omega_m\cap\Omega_{m+1}$，则 u_j 值始终不更

新，只有那些属于 Ω_{m+1}/Ω_m 的节点函数值才在迭代过程中得到改进。总之，无论网格节点采用什么编号规则，子空间迭代过程都是相同的。

全空间光滑是利用椭圆算子的光滑特性，使 Ω_m 层上的数值解 $u_m(\Omega_m)$ 具有一定的精度和光滑性，以保证外推工作的继续进行。全空间光滑要对 Ω_m 上所有的节点作迭代。

由此可见，子空间迭代和全空间光滑完全可以统一在同一个迭代子程序中，它们的计算公式是统一的，只是计算过程略有不同。不仿在子程序中设置一个哑元，以区分并灵活使用这两种不同的迭代过程。

子空间迭代和全空间光滑过程的精度控制是很重要的。在 I.S.S. 算法中各网格层上都进行了子空间迭代和全空间光滑，显然，若精度控制过严会浪费计算量，精度要求过低又不能获得足够精度的数值解。下面给出一种精度控制标准。

最终精度要求：

$$\frac{\|u-\tilde{u}_l\|}{\|\tilde{u}_l\|}\leqslant \varepsilon; \tag{5.18}$$

最终精度控制标准：

$$\frac{\|u_l-\tilde{u}_l\|}{\|\tilde{u}_l\|}\leqslant\frac{\varepsilon}{4},\quad \frac{\|\tilde{u}_{l-1}-\tilde{u}_l\|}{\|\tilde{u}_l\|}\leqslant\frac{\varepsilon}{4}; \tag{5.19}$$

在 $\Omega_m(3\leqslant m\leqslant l-1)$ 层上的全空间光滑终止控制：

$$\begin{cases}\dfrac{\|u_m-\tilde{u}_m\|}{\|\tilde{u}_m\|}\leqslant 2^{(l-m)\alpha}\dfrac{\varepsilon}{4},\\[2ex] \dfrac{\|\tilde{u}_{m-1}-\tilde{u}_m\|}{\|\tilde{u}_m\|}\leqslant 2^{(l-m)\alpha}\dfrac{\varepsilon}{4};\end{cases}\quad \alpha\text{满足(5.5)式,} \tag{5.20}$$

在 $\Omega_m(3\leqslant m\leqslant l-1)$层上的子空间迭代终止控制：

$$\|\tilde{u}_m\|\approx\|\tilde{u}_m\|. \tag{5.21}$$

说明 由(5.5)式，在节点P处，有

$$u_m(P)-u(P)=\phi(P)h_m^{\alpha}+O(h_m^{\alpha+1}),$$

其中 $\phi(P)$ 与 h 无关；若采用 Euclid 模，则上式为

$$\|u_m - u\| = \|\phi\| h_m^\alpha + O(h_m^{\alpha+1})。$$

从而有

$$|u_m(P) - u(P)| \sim \left(\frac{1}{2}\right)^\alpha |u_{m-1}(P) - u(P)|,$$

$$\|u_m - u\| \sim \left(\frac{1}{2}\right)^\alpha \|u_{m-1} - u\|。$$

因此，能推出

$$|u_m(P) - u(P)| \lesssim \frac{1}{2^\alpha - 1} |u_m(P) - u_{m-1}(P)|, \qquad (5.22)$$

$$\|u_m - u\| \lesssim \frac{1}{2^\alpha - 1} \|u_m - u_{m-1}\|。 \qquad (5.23)$$

(5.22)式和(5.23)式表明可以用相邻两网格层近似解的误差去估计近似解 u_m 与真解 u 的误差。

另一方面，由于

$$\|u - \tilde{u}_m\| \leqslant \|u - u_m\| + \|u_m - \tilde{u}_m\|,$$

故根据(5.23)式，有

$$\|u - \tilde{u}_m\| \lesssim \|u_{m-1} - u_m\| + \|u_m - \tilde{u}_m\|$$
$$\lesssim \|u_{m-1} - \tilde{u}_{m-1}\| + \|\tilde{u}_{m-1} - \tilde{u}_m\| + 2\|u_m - \tilde{u}_m\|,$$

即有

$$\frac{\|u - \tilde{u}_m\|}{\|\tilde{u}_m\|} \lesssim \frac{\|u_{m-1} - \tilde{u}_{m-1}\|}{\|\tilde{u}_{m-1}\|} + \frac{\|\tilde{u}_{m-1} - \tilde{u}_m\|}{\|\tilde{u}_m\|} + 2\frac{\|u_m - \tilde{u}_m\|}{\|\tilde{u}_m\|}。$$

由此可见，只要用(5.19)式作为迭代终止标准，便可保证最终的精度要求(5.18)式。

(5.19)式的第一式可用于衡量在第 l 网格层上全空间光滑的精度要求并作为光滑的结束标准。(5.19)式的第二式则反映了网格层细分程度足够与否，若(5.19)式的第二式不满足，还可将网格再细分一次，再作计算，以致满足规定精度后，网格细分就可不再进行，整个 I.S.S. 算法到此结束。(5.20) 式用于控制中间网格层上的全空间光滑的结束与否，它与(5.19)式相似，只是精

度要求放宽了 $2^{(l-m)\alpha}$ 倍。这是因为 $\Omega_m(m<l)$ 的网格参数 h_m 和 Ω_l 的网格参数 h_l 之间存在着 $h_m=2^{l-m}h_l$ 的关系的缘故；另外由(5.5)式有

$$\|u-u_{m-1}\|\approx 2^{\alpha}\|u-u_m\|,$$

由此推出

$$\|u-u_m\|\approx 2^{(l-m)\alpha}\|u-u_l\|; \tag{5.24}$$

这样，可以简单仿效(5.19)式给出(5.20)式。当然(5.20)式的控制精度还可以适当放宽些，这不仅是因为过渡层控制过严会浪费工作量，而且在逐层求解逐层控制的过程中可以弥补前面的某些不足之处。

当然也可以不用相对误差控制，而且直接误差控制还可以适当简化些，下面给出另一组**精度控制标准**。

最终精度要求：$\|u-\tilde{u}_l\|<\varepsilon$；

最终精度控制：$\|u_l-u_{l-1}\|\leqslant\varepsilon$； (5.25)

全空间光滑控制：

$$\|u_m-u_{m-1}\|\leqslant 2^{(l-m)\alpha}\varepsilon,\ \ 3\leqslant m\leqslant l-1;$$

子空间迭代控制：$\|\tilde{u}_m\|\approx\|\bar{u}\|$。

说明 这里 $\|\cdot\|$ 也同样可换成 $|\cdot|_\infty$。由(5.22)式和(5.23)式可知

$$\|u-\tilde{u}_l\|<\|u-u_l\|\lesssim\|u_l-u_{l-1}\|, \tag{5.26}$$

这就是说，若 $\|u_l-u_{l-1}\|\leqslant\varepsilon$，则最终输出的外推值的精度至少达到 ε，当然这种估计是保守的。(5.25)式实际上也是网格细分过程的终止控制标准。

I.S.S. 算法最简单直观的层间光滑精度控制办法是

$$\|u_m-u_{m-1}\|\leqslant h_m^{\alpha},\ \ 1\leqslant m\leqslant l. \tag{5.27}$$

这里 $\|\cdot\|$ 也可换成 $|\cdot|_\infty$。由(5.22)式和(5.23)式可知，这样可保证 $\|u-\overset{*}{u}_m\|\leqslant\frac{1}{2^{\alpha}-1}h_m^{\alpha}$ $(1\leqslant m\leqslant l)$；当然，最终外推值的精度要比(5.27)式的控制精度好得多。

应该强调，(5.22)式和(5.23)式是有意义的，它不仅在 I.S.S. 算法精度控制标准确定方面起着关键作用，而且使 I.S.S. 算法最终能输出数值解和真解(尽管真解并不知道)误差的保守估计，这种**事后误差估计**在实际应用中是非常有意义的。

§5 子空间迭代收敛性分析

I.S.S. 算法的核心问题之一是各网格层上子空间迭代的收敛性分析及最佳松弛因子ω的选择问题。本节借助于局部 Fourier 分析方法，得到了三个重要结论：

1. 子空间迭代的收敛因子h无关地小于1；
2. 对于各种含参数的松弛法，可确定出最佳松弛因子；
3. 对于不同迭代格式和不同类型的网格点，可采用变参数松弛。

5.1 统一说明

考虑定义在Ω_m网格层上的离散方程

$$L_m u_m = f_m.$$

无论采用何种网格分划，若改变网格节点的编号，则L_m的行和列必相应地作若干次同时互换，L_m的结构特点也将随之改变，但是这并不改变L_m的谱半径。引理 5.1 就说明了这一事实。

引理5.1 矩阵L_m的第i行和第i列与第j行和第j列同时互换后，L_m的谱半径不变。

证明 设$P(i,j)$为对L_m施行第i行和第j行对换的初等阵，易知第i列和第j列施行对换的初等阵也为$P(i,j)$，且L_m的第i,j行和列都同时互换后的矩阵为$P(i,j)L_mP(i,j)$，所以

$$\rho(P(i,j)L_mP(i,j))=\rho(P^2(i,j))\rho(L_m)=\rho(L_m),$$

其中$\rho(A)$表示矩阵A的谱半径。 □

根据引理5.1，各网格层上L_m的谱半径并不随网格节点编

号的改变而改变，因此在规则编号的情况下分析各层网格上离散算子的谱半径是具有普遍性的。本节统一采用坐标式编号规则，并采用第二章中的局部 Fourier 分析方法，分析各种迭代算子在子空间迭代中的收敛特性。

正如第二章§2中所述，对局部离散格式 $[S_k]$ 中的指标集 V 进行点集分解

$$V=V^+\cup V^0\cup V^-,$$

那么局部离散格式 $[S_k]$ 也可分解为

$$[S_k]=[S_k]^++[S_k]^0+[S_k]^-,$$

从而 $[S_k]$ 的特征值 $\lambda(\theta,h)$ 也相应分解为

$$\lambda(\theta,h)=\lambda^+(\theta,h)+\lambda^0(\theta,h)+\lambda^-(\theta,h),$$

这样局部迭代过程的特征值可由定理 2.2 确定，即

$$[S_k]\varphi(\theta,x)=A(h,\omega)\varphi(\theta,x),\quad |\theta|\leqslant\pi,$$

$$A(h,\omega)=\frac{(1-\omega)\lambda^0-\omega\lambda^-}{\lambda^0+\omega\lambda^+},\tag{5.28}$$

其中 $\varphi(\theta,x)$ 是 $[S_k]$ 的特征向量。

本节将对 5 点差分格式、9 点差分格式、直角三角线性元格式、矩形双线性元格式和一般三角线性元格式(这些格式参见第一章）施行 ω-Jacobi 松弛、SOR 方法、SOR 线松弛和 RB-Jacobi 松弛，并分析这些方法在子空间迭代过程中的收敛特性。

对于矩形分划情形，在子空间迭代中会出现三种不同类型的点，如图 5.7 所示。若统一用 9 点局部离散格式

$$\begin{pmatrix} s_{-1,1} & s_{0,1} & s_{1,1}\\ s_{-1,0} & s_{0,0} & s_{1,0}\\ s_{-1,-1} & s_{0,-1} & s_{1,-1}\end{pmatrix}$$

的形式描述前面所述诸格式，则这三类点所对应的局部离散格式分别为

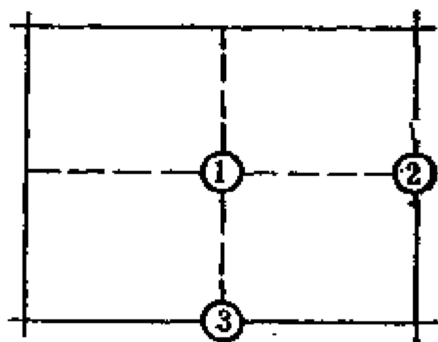


图 5.7

第一类点局部格式　　　　第二类点局部格式

$$\begin{pmatrix} 0 & s_{0,1} & 0 \\ s_{-1,0} & s_{0,0} & s_{1,0} \\ 0 & s_{0,-1} & 0 \end{pmatrix}, \quad \begin{pmatrix} s_{-1,1} & 0 & s_{1,1} \\ s_{-1,0} & s_{0,0} & s_{1,0} \\ s_{-1,-1} & 0 & s_{1,-1} \end{pmatrix},$$

第三类点局部格式

$$\begin{pmatrix} s_{-1,1} & s_{0,1} & s_{1,1} \\ 0 & s_{0,0} & 0 \\ s_{-1,-1} & s_{0,-1} & s_{1,-1} \end{pmatrix}.$$

因此子空间中迭代的收敛特性分析，就是针对各种不同的离散格式和松弛方法，利用（5.28）式对上述三类点的局部格式进行分析。下面将用$|A_j(\cdot)|$表示对第 j 类点($j=1,2,3$)施行局部迭代的特征值，且记 $\rho_j^*(\omega)=\sup\limits_{h\to 0}\{|A_j(h,\omega)|\}$。

为简单起见，本节只就 Poisson 方程这一模型问题的相应情形讨论之。

5.2　5点差分格式情形

这里讨论在正方形分划下的 5 点差分格式情形（参见（1.15）式）。此时，子空间迭代中三类不同点的局部格式如下：

第一类点局部格式　　第二类点局部格式　　第三类点局部格式

$$\begin{pmatrix} & -1 & \\ -1 & 4 & -1 \\ & -1 & \end{pmatrix}, \quad \begin{pmatrix} & 0 & \\ -1 & 4 & -1 \\ & 0 & \end{pmatrix}, \quad \begin{pmatrix} & -1 & \\ 0 & 4 & 0 \\ & -1 & \end{pmatrix}.$$

先考虑 ω-Jacobi 松弛($0<\omega\leqslant 1$)。

对于第一类点，有 $\lambda^0=4$，$\lambda^+=0$，$\lambda^-=-2(\cos\theta_1+\cos\theta_2)$，所以由(5.28)式，有

$$|A_1(h,\omega)|=1-\frac{\omega}{2}(2-\cos\theta_1-\cos\theta_2), \qquad \rho_1^*(\omega)=1.$$

对于第二类点，有 $\lambda^0=4$，$\lambda^+=0$，$\lambda^-=-2\cos\theta_1$，所以由(5.28)式有

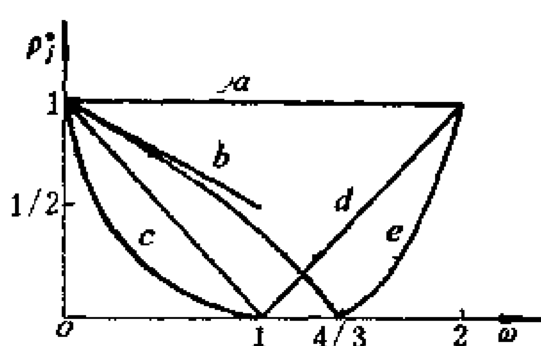
$$|A_2(h,\omega)|=1-\omega(2-\cos\theta_1)/2, \quad \rho_2^*(\omega)=1-\omega/2.$$

对第三类点，有 $\lambda^0=4$，$\lambda^+=0, \lambda^-=-2\cos\theta_2$，由(5.28)式得

$$|A_3(h,\omega)|=1-\omega(2-\cos\theta_2)/2,\quad \rho_3^*(\omega)=1-\omega/2.$$

表5.1 五点差分格式情形，子空间迭代收敛特性

松弛方法	各类点收敛特性
ω-Jacobi 松弛	$\rho_1^*=1,\quad \rho_2^*=1-\frac{\omega}{2},\quad \rho_3^*=1-\frac{\omega}{2}$
逐点SOR 方法	$\rho_1^*=1,\ \rho_2^*=\left\|\frac{4-3\omega}{4-\omega}\right\|,\ \rho_3^*=\left\|\frac{4-3\omega}{4-\omega}\right\|$
逐线SOR 方法	$\rho_1^*=1,\ \rho_2^*=\left\|\frac{4-3\omega}{4-\omega}\right\|,\ \rho_3^*=\|1-\omega\|$
RB-Jacobi 松弛	$\rho_j^*=(1-\omega)^2,\ j=1,2,3.$

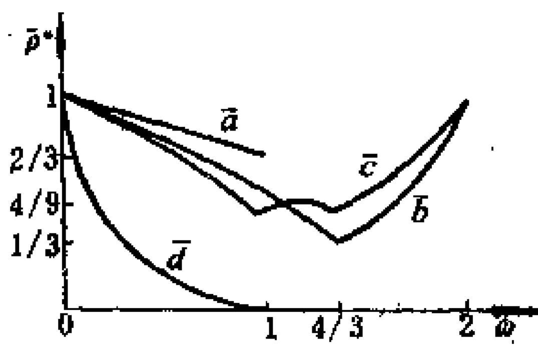


a. $\rho_j^*=1$,　b. $\rho_j^*=1-\frac{1}{2}\omega$,

c. $\rho_j^*=(1-\omega)^2$,　d. $\rho_j^*=|1-\omega|$.

e. $\rho_j^*=\left|\frac{4-3\omega}{4-\omega}\right|$.

图5.8



$\bar{a}$. ω-Jacobi 松弛,

$\bar{b}$. SOR 方法,

$\bar{c}$. SOR 线松弛,

$\bar{d}$. RB-Jacobi 松弛.

图5.9

再考虑逐点 SOR 方法和逐线 SOR 方法在子空间迭代中的收敛特性，仍用局部 Fourier 分析法讨论，其结果见表5.1.

RB-Jacobi 松弛法在子空间迭代中只有一种类型的点，其

$\lambda^0=4$，$\lambda^+=\lambda^-=0$，所以半个迭代步有收敛因子$(1-\omega)$，对完整迭代步有

$$A_j(h,\omega)=(1-\omega)^2,\ \rho_j^*(\omega)=(1-\omega)^2,\ j=1,2,3.$$

最后，把三类不同点局部格式在子空间迭代中的收敛特性表达式ρ_j^*列于表5.1中，将ρ_j^*的曲线、子空间迭代的平均收敛特性$\bar{\rho}^*=\frac{1}{3}\sum_{j=1}^{3}\rho_j^*$的曲线分别绘置于图5.8和图5.9中。

5.3 9点差分格式情形

这里仅讨论在正方形分划下的9点差分格式情形(参见(1.17)式)。此时，子空间迭代中三类不同点的局部格式如下：

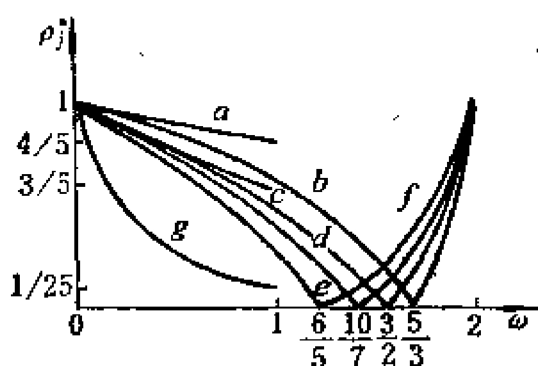
第一类点局部格式　　第二类点局部格式　　第三类点局部格式

$$\begin{pmatrix}0 & -4 & 0\\ -4 & 20 & -4\\ 0 & -4 & 0\end{pmatrix},\quad \begin{pmatrix}-1 & 0 & -1\\ -4 & 20 & -4\\ -1 & 0 & -1\end{pmatrix},\quad \begin{pmatrix}-1 & -4 & -1\\ 0 & 20 & 0\\ -1 & -4 & -1\end{pmatrix}.$$

仿照§5.2中的办法，将讨论的结果列于表5.2和图5.10、图5.12中。

表5.2　9点差分格式情形，子空间迭代收敛特性

松弛方法	各类点收敛特性
ω-Jacobi 松弛	$\rho_1^*=1-\frac{1}{5}\omega$，　$\rho_2^*=1-\frac{2}{5}\omega$，　$\rho_3^*=1-\frac{2}{5}\omega$
逐点SOR方法	$\rho_1^*=\left\lvert\frac{5-3\omega}{5-2\omega}\right\rvert$，　$\rho_2^*=\left\lvert\frac{10-7\omega}{10-3\omega}\right\rvert$，　$\rho_3^*=\left\lvert\frac{10-7\omega}{10-3\omega}\right\rvert$
逐线SOR方法	$\rho_1^*=\left\lvert\frac{3-2\omega}{3-\omega}\right\rvert$，　$\rho_2^*=\left\lvert\frac{10-7\omega}{10-3\omega}\right\rvert$，　$\rho_3^*=\left\lvert\frac{6-5\omega}{6-\omega}\right\rvert$
RB-Jacobi 松弛	$\rho_j^*=\left(1-\frac{4}{5}\omega\right)^2$，$j=1,2,3.$



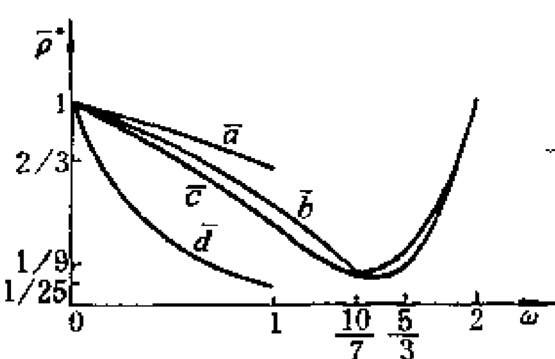
a. $\rho_j^* = 1 - \frac{1}{5}\omega$,　　b. $\rho_j^* = \left|\frac{5-3\omega}{5-2\omega}\right|$,

c. $\rho_j^* = 1 - \frac{2}{5}\omega$,　　d. $\rho_j^* = \left|\frac{3-2\omega}{3-\omega}\right|$,

e. $\rho_j^* = \rho\left|\frac{10-7\omega}{10-3\omega}\right|$,　　f. $\rho_j^* = \left|\frac{6-5\omega}{6-\omega}\right|$,

g. $\rho_j^* = \left(1 - \frac{4}{5}\omega\right)^2$.

图5.10



$\bar{a}$. ω-Jacobi松弛.

$\bar{b}$. SOR 方法.

$\bar{c}$. SOR线松弛.

$\bar{d}$. RB-Jacobi松弛.

图5.11

5.4 矩形双线性元格式情形

这里仅讨论正方形双线性元格式情形，参见(1.27b)式. 此时，子空间迭代中三类不同点的局部格式如下：

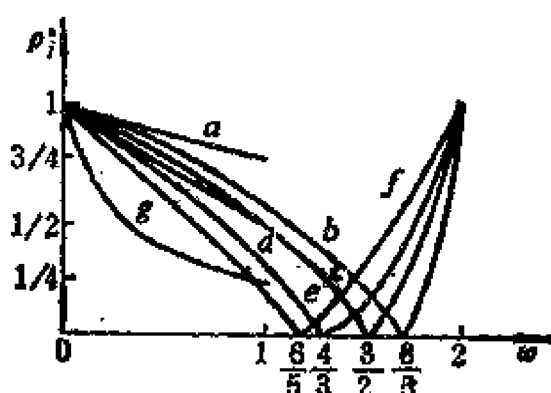
第一类点局部格式　　第二类点局部格式　　第三类点局部格式

$$\begin{pmatrix} 0 & -\frac{1}{3} & 0 \\ -\frac{1}{3} & \frac{8}{3} & -\frac{1}{3} \\ 0 & -\frac{1}{3} & 0 \end{pmatrix}, \quad \begin{pmatrix} -\frac{1}{3} & 0 & -\frac{1}{3} \\ -\frac{1}{3} & \frac{8}{3} & -\frac{1}{3} \\ -\frac{1}{3} & 0 & -\frac{1}{3} \end{pmatrix}. \quad \begin{pmatrix} -\frac{1}{3} & -\frac{1}{3} & -\frac{1}{3} \\ 0 & \frac{8}{3} & 0 \\ -\frac{1}{3} & -\frac{1}{3} & -\frac{1}{3} \end{pmatrix}.$$

仍然仿照§5.2中的办法，分别对这三类点的局部格式讨论之，并将结果列于表5.3和图5.12、图5.13中.

表5.3　正方形双线性元格式情形，子空间迭代收敛特性

松弛方法	各类点收敛特性
ω-Jacobi 松弛	$\rho_1^* = 1-\frac{\omega}{2}$，　$\rho_2^* = 1-\frac{\omega}{4}$，　$\rho_3^* = 1-\frac{\omega}{4}$
逐点SOR方法	$\rho_1^* = \left\lvert\frac{4-3\omega}{4-\omega}\right\rvert$，$\rho_2^* = \left\lvert\frac{8-5\omega}{8-3\omega}\right\rvert$，$\rho_3^* = \left\lvert\frac{8-5\omega}{8-3\omega}\right\rvert$
逐线SOR方法	$\rho_1^* = \left\lvert\frac{6-5\omega}{6-\omega}\right\rvert$，$\rho_2^* = \left\lvert\frac{8-5\omega}{8-3\omega}\right\rvert$，$\rho_3^* = \left\lvert\frac{3-2\omega}{3-\omega}\right\rvert$
RB-Jacobi 松弛	$\rho_j^* = \left(1-\frac{\omega}{2}\right)^2$



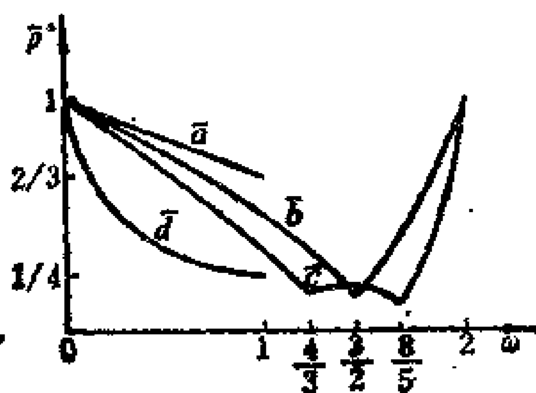
a. $\rho_j^* = 1-\frac{1}{4}\omega$.　b. $\rho_j^* = \left\lvert\frac{8-5\omega}{8-3\omega}\right\rvert$.

c. $\rho_j^* = \left\lvert\frac{3-2\omega}{3-\omega}\right\rvert$.　d. $\rho_j^* = 1-\frac{1}{2}\omega$.

e. $\rho_j^* = \left\lvert\frac{4-3\omega}{4-\omega}\right\rvert$.　f. $\rho_j^* = \left\lvert\frac{6-5\omega}{6-\omega}\right\rvert$.

g. $\rho_j^* = \left(1-\frac{1}{2}\omega\right)^2$.

图5.12



$\bar{a}$. ω-Jacobi 松弛.

$\bar{b}$. SOR线松弛.

$\bar{c}$. SOR方法.

$\bar{d}$. RB-Jacobi松弛.

图5.13

5.5　三角线性元格式情形

三角线性元格式由(1.21)式表示。子空间迭代中三类不同点的局部格式相应如下：

第一类点局部格式　第二类点局部格式　　第三类点局部格式

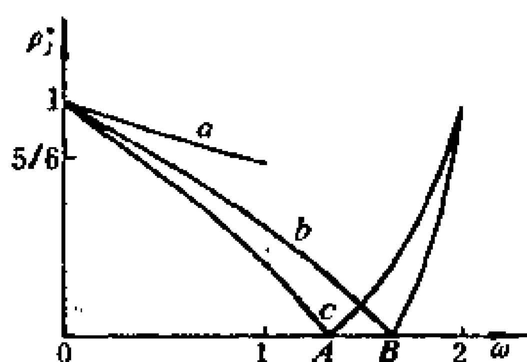
$$\begin{pmatrix} 0 & -c_2 & 0 \\ -c_3 & 2c & -c_3 \\ 0 & -c_2 & 0 \end{pmatrix}, \quad \begin{pmatrix} -c_1 & 0 & 0 \\ -c_3 & 2c & -c_3 \\ 0 & 0 & -c_1 \end{pmatrix}, \quad \begin{pmatrix} -c_1 & -c_2 & 0 \\ 0 & 2c & 0 \\ 0 & -c_2 & -c_1 \end{pmatrix},$$

其中，$c_i = \operatorname{ctg}\alpha_i (i=1,2,3)$，$c = c_1 + c_2 + c_3$，$\alpha_i (i=1,2,3)$是三角单元的三个角度参数，参见图1.7。

分别对这三类局部格式讨论前述几种松弛法的子空间迭代收敛特性，将结果列于表5.4和图5.14、图5.15中。

表5.4　三角线性元格式情形，子空间迭代收敛特性

松弛方法	ρ_j^*，一般三角形单元	ρ_j^*，等边三角形单元	ρ_j^*，等腰直角三角形单元
ω-Jacobi松弛	$\rho_1^* = 1 - \frac{c_1}{c}\omega$	$\rho_1^* = 1 - \frac{\omega}{3}$	$\rho_1^* = 1$
	$\rho_2^* = 1 - \frac{c_2}{c}\omega$	$\rho_2^* = 1 - \frac{\omega}{3}$	$\rho_2^* = 1 - \frac{\omega}{2}$
	$\rho_3^* = 1 - \frac{c_3}{c}\omega$	$\rho_3^* = 1 - \frac{\omega}{3}$	$\rho_3^* = 1 - \frac{\omega}{2}$
逐点SOR方法	$\rho_1^* = \left\lvert 1 - \frac{2c_1\omega}{2c - \omega(c_2 + c_3)} \right\rvert$	$\rho_1^* = \left\lvert 1 - \frac{\omega}{3-\omega} \right\rvert$	$\rho_1^* = 1$
	$\rho_2^* = \left\lvert 1 - \frac{2c_2\omega}{2c - \omega(c_1 + c_3)} \right\rvert$	$\rho_2^* = \left\lvert 1 - \frac{\omega}{3-\omega} \right\rvert$	$\rho_2^* = \left\lvert 1 - \frac{2\omega}{4-\omega} \right\rvert$
	$\rho_3^* = \left\lvert 1 - \frac{2c_3\omega}{2c - \omega(c_1 + c_2)} \right\rvert$	$\rho_3^* = \left\lvert 1 - \frac{\omega}{3-\omega} \right\rvert$	$\rho_3^* = \left\lvert 1 - \frac{2\omega}{4-\omega} \right\rvert$
逐线SOR方法	$\rho_1^* = \left\lvert 1 - \frac{2c_1\omega}{2(c_1 + c_3) - \omega c_3} \right\rvert$	$\rho_1^* = \left\lvert 1 - \frac{2\omega}{2-\omega} \right\rvert$	$\rho_1^* = 1$
	$\rho_2^* = \left\lvert 1 - \frac{2c_2\omega}{2c - \omega(c_1 + c_3)} \right\rvert$	$\rho_2^* = \left\lvert 1 - \frac{\omega}{3-\omega} \right\rvert$	$\rho_2^* = \left\lvert 1 - \frac{2\omega}{4-\omega} \right\rvert$
	$\rho_3^* = \left\lvert 1 - \frac{2c_3\omega}{2(c_1 + c_3) - \omega c_1} \right\rvert$	$\rho_3^* = \left\lvert 1 - \frac{2\omega}{2-\omega} \right\rvert$	$\rho_3^* = \lvert 1 - \omega \rvert$



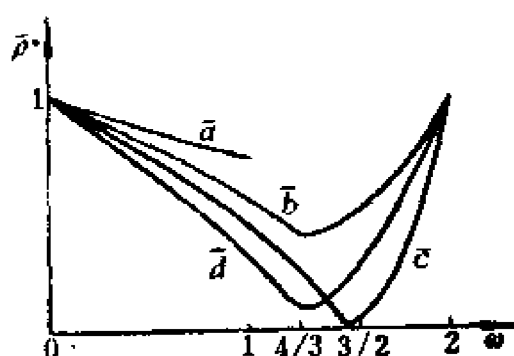
a. $\rho_j^* = 1 - \frac{c_1}{c}\omega$,

b. $\rho_j^* = \left|1 - \frac{2c_1\omega}{2c - \omega(c_2+c_3)}\right|$,

c. $\rho_j^* = \left|1 - \frac{2c_1\omega}{2(c_1+c_3) - \omega c_3}\right|$.

$A = \frac{2(c_1+c_3)}{2c_1+c_3}$, $B = \frac{2c}{c+c}$

图 5.14



$\bar{a}$. ω-Jacobi 松驰(等边△),

$\bar{b}$. SOR 方法(等腰直角△),

$\bar{c}$. SOR 方法(等边△),

$\bar{d}$. GS线松驰(等边△).

图 5.15

5.6 综合分析

正如第三章指出的，逐点迭代方法在实施过程中，迭代前后近似解数值的改变只是在逐个网格点及其邻点(离散格式作用于某个网格点时所涉及到的网格点)上发生，局部 Fourier 分析法反应了迭代的局部性质，因此，可用局部 Fourier分析方法分别独立地分析周期分布的第一、二、三类点的收敛特性。另外，由于在实际迭代过程中，迭代值的改进是局部地、顺次地在第一、二、三类点及其邻点上反复进行的，所以可以认为总迭代效果是这三类点处迭代效果的平均，即取 $\bar{\rho}^* = \frac{1}{3}\sum_{j=1}^{3}\rho_j^*$ 表示某种离散格式采用某种松弛方法在子空间迭代过程中所表现出来的平均效应，参见图5.9、图 5.11、图5.13和图 5.15，这种子空间迭代平均收敛效果也在数值实验中得到了证实。

综合分析§5.2～§5.5中的图和表，即可得到子空间迭代中普

遍成立的四个结论。

第一，无论网格层的细密程度如何，在各网格层上子空间迭代的收敛因子相同且 h 无关地小于 1。这一结论表明，当网格无限细分时，子空间迭代仍保持极好的收敛速度。这是不难理解的，因为网格剖分是逐层加密的，任一网格层子空间中的网格点相对于次层粗网格点的分布规则是完全一样的，所以子空间迭代在各网格层次上的收敛因子是相同的。

第二，对于各种带松弛参数的迭代方法，可确定出各网层上子空间迭代都适用的最佳松弛因子。这在图5.9、图5.11、图5.13和图5.15中已表示得非常明确。例如，不论对 5 点差分格式、直角三角线性元格式还是正方形双线性元格式，当采用逐点SOR方法作子空间迭代时，最佳松弛因子都是 4/3，这与数值计算效果相符合。选用这些最佳松弛因子，可以大大地加快子空间迭代的收敛速度。

第三，对不同类型的点使用不同最佳松弛因子的松弛步，这对进一步加快子空间迭代的收敛速度无疑是有益的。

第四，在用 I.S.S. 算法数值求解Poisson方程时，根据本节中的诸表和图不难看出，各种松弛方法用于子空间迭代的优势顺序大致是：

ω-Jacobi 松弛<逐点SOR方法<逐线SOR方法<RB松弛，其中，“<”表示“不优于”之意。

§6 工作量估计与算例比较

设 $m\geqslant 3$，且用 $W_{外}(\Omega_m)$、$W_{内}(\Omega_m)$、$W(\Omega_m/\Omega_{m-1})$、$W(\Omega_m)$ 和 W_m 分别表示在 Ω_m 网格层上进行外推过程、内插过程、子空间迭代过程、全空间光滑过程的计算量和所有这些计算量的总和。易知 $W_{外}(\Omega_m)\approx W_{内}(\Omega_m)=O(N_m)$；由于子空间迭代的 h 无关收敛特性，故知有限步子空间迭代所需计算量 $W(\Omega_m/\Omega_{m-1})=$

$O(N_m)$；因为在 Ω_m 网格层上只作有限步整体光滑迭代，所以，$W(\Omega_m)=O(N_m)$；总之，在 Ω_m 网格层上所花费的全部计算量是 $W_m=O(N_m)$。

在上述分析的基础上，很容易获得 I.S.S. 算法数值求解 $L_l u_l = f_l$ 的总计算量估计：

$$
\begin{aligned}
W &= W_l + W_{l-1} + \cdots + W_2 + W_1 \\
&= O(N_l) + O(N_{l-1}) + \cdots + O(N_2^2) + O(N_1^2) \\
&\approx O(N_l)\left(1+\left(\frac{1}{4}\right)^1+\left(\frac{1}{4}\right)^2+\cdots+\left(\frac{1}{4}\right)^{l-1}\right)=O(N_l),\quad l\geqslant 3.
\end{aligned}
$$

为了观察 I.S.S.算法的实际效果，下面给出两个算例的数值比较。

算例1 定解问题

$$
\begin{cases}
-\Delta u=\pi^2\sin\pi x(5\cos 2\pi y-1), & \Omega=(0,1)\times(0,1),\\
u|_{\partial\Omega}=0
\end{cases}
$$

的准确解为 $u(x,y)=\sin\pi x(\cos 2\pi y-1)$。

将[0,1]×[0,1]按§2中办法分半细分数次，直至 Ω_4 具有289个网格节点。各网格层上的全空间光滑采用带有该层最佳松弛因子的 SOR 方法。各网格层子空间迭代一律采用 $\omega=4/3$ 的 SOR 方法。表 5.5 中列出了 I.S.S. 算法在执行过程中的有关数据。表

表5.5 矩形双线性元I.S.S.算法的计算数据

网格层次	节点总数	ω_{opt}	控制精度	子空间迭代次数	全空间光滑次数	折算到 Ω_4 上的迭代次数
1	9	1.25	0.353553		2	0.1
2	25	1.45	0.044194		7	0.5
8	81	1.55	0.005524	10	9	4.1
4	289	1.65	0.000696	22	8	24.5

注：子空间迭代中采用 $\omega=4/3$，若 $\omega=8/5$ 效果略好一点。

5.6中列出了采用I.S.S.算法与SOR方法的计算量比较。这些有关数据虽然是就矩形双线性元情形的计算和比较结果，但它相仿于用三角线性元或5点差分格式的计算结果。正如表5.6中的

表5.6 I.S.S.算法与SOR方法计算效果比较

网格层	节点数	数值方法	$\lvert u-u_h\rvert_\infty$	迭代次数	备注
Ω_4	289	I.S.S.	$0.38E-3$	30	已折算
Ω_4	289	SOR	$0.38E-2$	41	$\omega_{opt}=1.65$
Ω_5	1089	SOR	$0.39E-3$	46	$\omega_{opt}=1.70$ 相当于Ω_4上180多次迭代

数据所示，在Ω_4网格层上采用I.S.S.算法，只需用相当于Ω_4层上的30次SOR迭代的计算量，即可获得具有0.38×10^{-3}精度的数值解；然而用带有最佳松弛因子的SOR方法计算之，需要Ω_4网格层上的41次迭代，也只能得到具有0.38×10^{-2}的精度的数值解，且在Ω_4层上用SOR方法不能达到0.38×10^{-3}这个精度；如果在Ω_5层上用带有最佳松弛因子的SOR方法，需要46次迭代步的计算量，此时数值解的精度虽可达精度0.38×10^{-3}，但其计算量相当于在Ω_4网格层上的180多次迭代步的计算量。另外，若对此算例在$\Omega_1\sim\Omega_4$网格层上做多层网格方法（V循环或W循环）计算，也可看到使用多层网格方法的计算量要比使用I.S.S.算大得多，而且数值解精度与使用SOR方法的精度相同。总之，I.S.S.算法的高效率高精度特征是明显的。

算例2 定解问题

$$\begin{cases}a(x,y)u_{xx}+b(x,y)u_{yy}=-f(x,y), & (x,y)\in\Omega,\\ u=g(x,y), & (x,y)\in\partial\Omega,\end{cases}$$

其中，

$$a(x,y)=x,\quad b(x,y)=xy,\quad f(x,y)=-6x^2(1+y^2),\quad \Omega=[0,1]\times[0,1],\qquad g(x,y)=\begin{cases}0, & x=0,\\ x^3, & y=0,\\ 1+y^3, & x=1,\\ x+x^3, & y=1.\end{cases}$$

真解是 $u(x,y)=x(x^2+y^3)$.

计算中 $\bar{\Omega}$ 被分成 $2^M\times 2^N$ 等分。分别就 $M=N=3$ 和 $M=N=4$ 的网格分划情形，用 I.S.S. 方法和其它经典迭代方法对本算例计算，并把计算结果列于表5.7a)、5.7b)和5.8a)、5.8b)中。这些数表中，工作量表示折算成细网格层上的迭代步数；精度采用 $|u-u_h|_\infty$；表中空白表示因数值计算不够稳定而未列出结果；每个表格下面 4 行表示为了达到 I.S.S. 方法所获得的精度而进行计算的数据；表格中带"*"标记的数据，表示在同一网格层计算无法或难以达到 I.S.S. 方法所获得的数值解精度，于是将网格再细分一次计算并将计算结果折换为原网格层上的计算量；表格中带有"**"记号的数据，表示在同一网格层上无法达到 I.S.S. 方法所获得的数值解精度，再在细分一次的网格层上计算，虽花费巨大工作量仍然未能达到所要求的精度，于是列出其所能达到的精度且把计算也折换为该网格层上的计算量。总之，从表 5.7a)～5.8b) 中的数值比较也可看出 I.S.S. 算法的高精度高效率特征。

作为本章结束，我们指出，I.S.S. 方法是基于渐近展式(5.5)作外推的，当数值解满足相容逼近精度 $\|u_m-u\|=O(h_m^2)$ 时，外推效果才显著，为此在 Ω_m 网格层上全空间光滑花费了一定的计算量。人们若能确定出松弛迭代过程的外推公式，则 I.S.S.方法中的全空间光滑计算量将能极大地被节省。

另外，只要与非线性问题 $Lu=f$ 相应的离散方程 $L_hu_h=f_h$ 的数值解 u_h 具有形如 (5.5) 式的渐近展开式；只要与网格层序列

表5.7a) I.S.S. 方法与其它数值方法计算效果比较，$M=N=3$

		矩形双线性元			三角线性元		
		SOR	Jacobi	GS	SOR	Jacobi	GS
逐点松弛	精度	.120BE-1		.12866E-1	.13120E-3	.23630E-2	.13505E-2
	工作量	18		28	16	66	39
RB-松弛	精度	.1212OE-1	.13210E-1	.12890E-1	.15855E-3	.15640E-2	.13350E-2
	工作量	14	32	28	13	36	37
I.S.S. 方法	精度	.39715E-2		.42200E-2	.95069E-4	.84084E-3	.50509E-3
	工作量	13		15	9	32	16
RB-松弛	精度	.39715E-2		.42200E-2	.95069E-4	.84084E-3	.50509E-3
	工作量	72*		428*	56*	160*	176*
逐点松弛	精度	.39715E-2		.42200E-2	.95069E-4	.84084E-3	.50509E-3
	工作量	168*		428*	68*	316*	184*

表5.7b) I.S.S. 方法与其它数值方法计算效果比较，$M=N=3$

		5点差分格式			9点差分格式		
		SOR	Jacobi	GS	SOR	Jacobi	GS
逐点松弛	精度	.13120E-3	.23630E-2	.13505E-2	.59907E-2	.81866E-2	.70120E-2
	工作量	16	66	39	16	55	34
RB-松弛	精度	.15855E-3	.15640E-2	.13350E-2	.60132E-2	.71951E-2	.68859E-2
	工作量	13	36	37	12	34	33
I.S.S.	精度	.95069E-4	.84084E-3	.50509E-3	.20015E-2	.26467E-2	.26771E-2
	工作量	9	32	16	13	25	15
RB-松弛	精度	.95069E-4	.84084E-3	.50509E-3	.20015E-2		.26771E-2
	工作量	56*	160*	176*	84*		524*
逐点松弛	精度	.95069E-4	.84084E-3	.50509E-3	.20015E-2		.26771E-2
	工作量	68*	316*	184*	96*		536*

表5.8e) I.S.S. 方法与其它数值方法计算结果比较，$M=N=4$

		矩形双线性元			三角线性元		
		SOR	Jacobi	GS	SOR	Jacobi	GS
逐点松弛	精度	.31648E-2		.34007E-2	.26051E-3	.70435E-3	.37229E-3
	工作量	66		132	39	334	189
RB-松弛	精度	.31735E-2	.34704E-2	.34033E-2	.22411E-3	.38457E-3	.38457E-3
	工作量	34	158	133	32	183	183
I.S.S.	精度	.69724E-3		.07135E-3	.21436E-4	.40889E-4	.12559E-3
	工作量	36		41	18	99	29
RB-松弛	精度	.81789E-3		.81819E-3	.21436E-4	.40889E-4	.12559E-3
	工作量	4004**		4004**	132*	960*	848*
逐点松弛	精度	.81813E-3		.81778E-3	.21436E-4	.40889E-4	.12559E-3
	工作量	4004**		4004**	160*	1916*	868*

表5.8 b) I.S.S. 方法与其它数值方法计算效果比较, $M=N=4$

		5点差分格式			9点差分格式		
		SOR	Jacobi	GS	SOR	Jacobi	GS
逐点松弛	精度	.26051E-3	.70435E-3	.37229E-3	.15829E-2		.18793E-2
	工作量	39	334	189	38		161
RB-松弛	精度	.22411E-3	.38457E-3	.38457E-3	.15855E-2	.19168E-2	.18824E-2
	工作量	32	183	183	31	170	158
I.S.S.	精度	.21436E-4	.40889E-4	.12559E-3	.36750E-3		.41038E-3
	工作量	18	99	29	30		44
RB-松弛	精度	.21436E-4	.40889E-4	.12559E-3	.40996E-3		.41290E-3
	工作量	132*	960*	848*	4004**		4004**
逐点松弛	精度	.21436E 4	.40889E-4	.12559E-3	.41002E-3		.41223E-3
	工作量	160*	1916*	868*	4004**		4004**

$\{\Omega_l\}$对应存在着一串非线性离散方程 (5.2)；那么用 I.S.S. 方法数值求解最细网格层 Ω_l 上的数值解就是可行的.下面介绍两种办法。

方法一，逐层推进地求解各网格层上相应非线性离散方程的数值解，并用外推和内插等手段获得较细网格层上的数值解的好初值，以加速逐层推进的过程。该方法可简述如下：

$$
\begin{cases}
\text{求解 } L_i u_i = f_i,\ \ i = 1,2, \\
\text{对 k 从 3 开始到 } l \text{ 结束，作循环计算，} \\
u_{k-1}(\Omega_{k-1}) = \mathrm{EXT}(u_{k-2},\ u_{k-1}), \\
\bar{u}_k(\Omega_k) = \mathrm{INT}(u_{k-1}(\Omega_{k-1})), \\
u_k(\Omega_k) = \mathrm{relax}(L_k, f_k, \bar{u}_k).
\end{cases}
$$

这里 EXT 和 INT 分别表示外推和内插过程，relax 则表示对非线性问题适用的子空间迭代和全空间光滑的迭代过程。

方法二，对 Ω_l 层上任给的近似值 $\bar{u}_l$，用逐层推进的办法求校正量 $\bar{v}_l$，同样可利用外推和内插过程加速推进过程。该方法可描述如下：

$$
\begin{cases}
\text{任给 } \bar{u}_l \text{ 计算} \\
\qquad d_i = I_i^l d_l = I_i^l (f_l - L_l \bar{u}_l),\ \ i = 1,2,\cdots,l, \\
\text{求解 } L_i \bar{v}_i = d_i,\ \ i = 1,2, \\
\text{对 k 从 3 开始到 } l \text{ 结束，作循环计算，} \\
\bar{v}_{k-1}(\Omega_{k-1}) = \mathrm{EXT}(\bar{v}_{k-2},\ \bar{v}_{k-1}), \\
v_k(\Omega_k) = \mathrm{INT}(\bar{v}_{k-1}(\Omega_{k-1})), \\
\bar{v}_k(\Omega_k) = \mathrm{relax}(L_i, d_i, v_k(\Omega_k)), \\
u_l = \bar{u}_l + \bar{v}_l.
\end{cases}
$$

特别地，方法二中若去掉外推过程，则方法二就相仿于多层网格方法的一个循环步，因此，根据多层网格方法的收敛特性，可推知此时的 I.S.S. 方法是与 h 无关收敛的。

第六章　解有限元方程的逐层分裂迭代法

考虑模型问题

$$\begin{cases}-\nabla(p\nabla u)+qu=f, & (x,\ y)\in\Omega,\\ u=\alpha, & (x,\ y)\in\partial\Omega,\end{cases}\tag{6.1}$$

与之相应的有限元方程是

$$a(u_h,\ v_h)=(f,\ v_h),\qquad \forall v_h\in V_h(\Omega_h),\tag{6.2}$$

其中，Ω_h 是区域 Ω 被剖分的结果，如无特殊说明 $V_h(\Omega_h)$ 就简记为 V_h，(6.1)和(6.2)式的有关记号和意义都与第二篇第一、二章的相同。本章指出，在有限元空间 V_h 的某些直和分解设计下，双线性泛函 $a(\cdot,\ \cdot)$ 满足强 Schwarz 不等式。在此基础上，本章讨论解有限元方程的分裂算法，还讨论它与多层网格方法、外推技术和超收敛性技术相结合的逐层分裂迭代法。

在本章中，总是假设 $a(\cdot,\ \cdot)$ 对称正定，并记 $\|\cdot\|^2=a(\cdot,\ \cdot)$。

§1　强 Schwarz 不等式

大家知道，(6.2)式中双线性泛函 $a(\cdot,\ \cdot)$ 的有界性成立，且

$$0<|a(u_h,v_h)|\leqslant\tilde{\gamma}a^{1/2}(u_h,u_h)a^{1/2}(v_h,v_h),\qquad \forall u_h,v_h\in V_h,$$

其中

$$\tilde{\gamma}=\sup\{|a(u_h,\ v_h)|/\|u_h\|\|v_h\|\},\ \ \forall u_h,\ \ v_h\in V_h.$$

可以证明，当 u_h 和 v_h 同属于有限元空间 V_h 时 $\tilde{\gamma}=1$。因为 $\|\cdot\|$ 是能量模，所以它符合 Schwartz 不等式，即有 $|a(u_h,v_h)|\leqslant\|u_h\|\|v_h\|$，故 $\tilde{\gamma}=1$。再定义

$$\bar{\bar{\gamma}}=\sup_{u_h,v_h\in V_h}\{|2a(u_h,v_h)|/(\|u_h\|^2+\|v_h\|^2)\},$$

一方面，由 $2\|u_h\|\|v_h\| \leqslant \|u_h\|^2 + \|v_h\|^2$ 和 $\tilde{\gamma}$ 及 $\tilde{\tilde{\gamma}}$ 的定义，有 $\tilde{\gamma} \geqslant \tilde{\tilde{\gamma}}$；另一方面，由

$$
\begin{aligned}
\tilde{\tilde{\gamma}} &= \sup\{|2a(u_h, v_h)|/(\|u_h\|^2 + \|v_h\|^2)\} \\
&= \sup\{|2a(u_h, v_h)|/(\|u_h\|^2 + \|v_h\|^2);\ \|u_h\| = \|v_h\| = 1\} \\
&\geqslant \sup\{|a(u_h, v_h)|;\ \|u_h\| = \|v_h\| = 1\} \\
&= \sup\{|a(u_h, v_h)|/\|u_h\|\,\|v_h\|;\ \|u_h\| = \|v_h\| = 1\} = \tilde{\gamma}
\end{aligned}
$$

又有 $\tilde{\tilde{\gamma}} \geqslant \tilde{\gamma}$。总之有 $\tilde{\gamma} = \tilde{\tilde{\gamma}} = 1$。

但是，如果有限元空间 V_h 可分解为两个非平凡子空间的直和($V_h = V_{\mathrm{I}} \oplus V_{\mathrm{II}}$)，则可能会有更强的 Schwarz 不等式

$$
\begin{cases}
|a(v_{\mathrm{I}}, v_{\mathrm{II}})| \leqslant \gamma \|v_{\mathrm{I}}\|\,\|v_{\mathrm{II}}\|, & \forall v_{\mathrm{I}} \in V_{\mathrm{I}},\ \forall v_{\mathrm{II}} \in V_{\mathrm{II}}, \\
0 < \gamma = \sup\{|a(v_{\mathrm{I}}, v_{\mathrm{II}})|/\|v_{\mathrm{I}}\|\,\|v_{\mathrm{II}}\|;\ v_{\mathrm{I}} \in V_{\mathrm{I}},\ v_{\mathrm{II}} \in V_{\mathrm{II}}\} < 1.
\end{cases} \tag{6.3}
$$

若将 $a(\cdot, \cdot)$ 看作内积，内积公式

$$
a(v_{\mathrm{I}}, v_{\mathrm{II}}) = \cos\alpha \|v_{\mathrm{I}}\|\,\|v_{\mathrm{II}}\|, \quad \forall v_{\mathrm{I}} \in V_{\mathrm{I}}, \quad \forall v_{\mathrm{II}} \in V_{\mathrm{II}}
$$

中的 α 是矢量 v_{I} 和 v_{II} 的夹角，则当 v_{I} 和 v_{II} 同属一个有限元空间 V_h 时，v_{I} 和 v_{II} 两矢量的夹角可能是 0 或 π，所以 $\gamma = 1$；只有当 V_h 能直和分解且内积的核只属于某一个子空间时，两空间中所有矢量的夹角才可能适合 $|\cos\alpha| = \gamma < 1$。

我们称(6.3)式中的 γ 为**收缩数**。由收缩数的定义可知，γ 只可能与基函数和网格分划及微分方程的系数有关。

读者将在§4中看到满足强 Schwarz 不等式的有限元空间的具体设计，还将在§2和§3中看到强 Schwarz 不等式在分裂算法和逐层分裂算法中的具体应用。

§2 分裂算法

考虑数值求解有限元方程（6.2)。设采用三角单元剖分；所有三角单元的顶点先编号，编号为 1 到 N_{I}；单元边中点后编号，编号为 $N_{\mathrm{I}}+1$ 到 N_{II}；设在单元顶点处和单元边中点处分别设置

基函数 φ_i 和 ψ_j，且有

$$V_{\text{I}} = \text{span}\{\varphi_i | i = 1, 2, \cdots, N_{\text{I}}\},$$

$$V_{\text{II}} = \text{span}\{\psi_j | j = N_{\text{I}} + 1, \cdots, N_{\text{II}}\},$$

$$V_h = V_{\text{I}} \oplus V_{\text{II}}.$$

此时，有限元方程(6.2)的系数矩阵有块状结构，(6.2)式可改写为

$$a(u_h, v_h) = \begin{pmatrix} A_h & C_h \\ C_h^T & B_h \end{pmatrix} \begin{pmatrix} U_{\text{I}} \\ U_{\text{II}} \end{pmatrix} = \begin{pmatrix} F_{\text{I}} \\ F_{\text{II}} \end{pmatrix}, \tag{6.4}$$

其中，$A_h = (a(\varphi_i, \varphi_j))$，$B_h = (a(\psi_k, \psi_l))$，$F_{\text{I}} = ((f, \varphi_i))$，$F_{\text{II}} = ((f, \psi_k))$，$i, j = 1, 2, \cdots, N_{\text{I}}$，$k, l = N_{\text{I}} + 1, \cdots, N_{\text{II}}$，$U_{\text{I}}$ 和 U_{II} 分别是对应于 u_h 在单元顶点处函数值和在单元边中点处函数值所构成的系数矢量。

利用(6.3)和(6.4)式，可以设计三种求解有限元方程的分裂算法。

定理6.1 若(6.3)式成立，则解有限元方程(6.2)或(6.4)的分裂算法一：

$$\begin{cases} a(u_{\text{I}}^{k+1}, v_{\text{I}}) = (f, v_{\text{I}}) - a(u_{\text{II}}^k, v_{\text{I}}), \forall v_{\text{I}} \in V_{\text{I}}, \\ a(u_{\text{II}}^{k+1}, v_{\text{II}}) = (f, v_{\text{II}}) - a(u_{\text{I}}^k, v_{\text{II}}), \forall v_{\text{II}} \in V_{\text{II}}, \end{cases} \tag{6.5}$$

或

$$\begin{cases} A_h U_{\text{I}}^{k+1} = F_{\text{I}} - C_h U_{\text{II}}^k, \\ B_h U_{\text{II}}^{k+1} = F_{\text{II}} - C_h^T U_{\text{I}}^k, \end{cases} \tag{6.6}$$

具有与 h 无关的收敛因子 γ。

证明 因为 $V_h = V_{\text{I}} \oplus V_{\text{II}}$ 且 u_{I}、$v_{\text{I}} \in V_{\text{I}}$，$u_{\text{II}}$、$v_{\text{II}} \in V_{\text{II}}$，所以由(6.4)式知，(6.2)可改写为

$$\begin{cases} a(u_{\text{I}}, v_{\text{I}}) = (f, v_{\text{I}}) - a(u_{\text{II}}, v_{\text{I}}), \quad \forall v_{\text{I}} \in V_{\text{I}}, \\ a(u_{\text{II}}, v_{\text{II}}) = (f, v_{\text{II}}) - a(u_{\text{I}}, v_{\text{II}}), \quad \forall v_{\text{II}} \in V_{\text{II}}. \end{cases} \tag{6.7}$$

(6.5)式和(6.7)式相减，得

$$a(u_{\text{I}}^{k+1} - u_{\text{I}}, v_{\text{I}}) = -a(u_{\text{II}}^k - u_{\text{II}}, v_{\text{I}}), \quad \forall v_{\text{I}} \in V_{\text{I}},$$

$$a(u_{\text{II}}^{k+1} - u_{\text{II}}, v_{\text{II}}) = -a(u_{\text{I}}^k - u_{\text{I}}, v_{\text{II}}), \quad \forall v_{\text{II}} \in V_{\text{II}}.$$

令 $v_{\text{I}} = u_{\text{I}}^{k+1} - u_{\text{I}}$，$v_{\text{II}} = u_{\text{II}}^{k+1} - u_{\text{II}}$，且利用强 Schwarz 不等式(6.3)，则有

$$\|u_{\mathrm{I}}^{k+1}-u_{\mathrm{I}}\|^2=|a(u_{\mathrm{II}}^k-u_{\mathrm{II}},\ u_{\mathrm{I}}^{k+1}-u_{\mathrm{I}})|$$
$$\leqslant\gamma\|u_{\mathrm{II}}^k-u_{\mathrm{II}}\|\cdot\|u_{\mathrm{I}}^{k+1}-u_{\mathrm{I}}\|,$$
$$\|u_{\mathrm{II}}^{k+1}-u_{\mathrm{II}}\|^2=|a(u_{\mathrm{I}}^k-u_{\mathrm{I}},\ u_{\mathrm{II}}^{k+1}-u_{\mathrm{II}})|$$
$$\leqslant\gamma\|u_{\mathrm{I}}^k-u_{\mathrm{I}}\|\|u_{\mathrm{II}}^{k+1}-u_{\mathrm{II}}\|,$$

所以有

$$\|u_{\mathrm{I}}^{k+1}-u_{\mathrm{I}}\|\leqslant\gamma^2\|u_{\mathrm{I}}^{k-1}-u_{\mathrm{I}}\|,$$
$$\|u_{\mathrm{II}}^{k+1}-u_{\mathrm{II}}\|\leqslant\gamma^2\|u_{\mathrm{II}}^{k-1}-u_{\mathrm{II}}\|。\square$$

定理6.2 若(6.3)式成立，则解有限元方程(6.2)或(6.4)的分裂算法二：

$$\begin{cases}a(u_{\mathrm{I}}^{k+1},v_{\mathrm{I}})=(f,v_{\mathrm{I}})-a(u_{\mathrm{II}}^k,v_{\mathrm{I}}), & \forall v_{\mathrm{I}}\in V_{\mathrm{I}},\\ a(u_{\mathrm{II}}^{k+1},v_{\mathrm{II}})=(f,v_{\mathrm{II}})-a(u_{\mathrm{I}}^{k+1},v_{\mathrm{II}}), & \forall v_{\mathrm{II}}\in V_{\mathrm{II}},\end{cases}\tag{6.8}$$

或

$$\begin{cases}A_hU_{\mathrm{I}}^{k+1}=F_{\mathrm{I}}-C_hU_{\mathrm{II}}^k,\\ B_hU_{\mathrm{II}}^{k+1}=F_{\mathrm{II}}-C_h^TU_{\mathrm{I}}^{k+1},\end{cases}\tag{6.9}$$

具有 h 无关的收敛因子 γ^2。

证明 仿定理(6.1)的证明，利用(6.7)和(6.8)式相减，有

$$a(u_{\mathrm{I}}^{k+1}-u_{\mathrm{I}},v_{\mathrm{I}})=-a(u_{\mathrm{II}}^k-u_{\mathrm{II}},v_{\mathrm{I}})=a(u_{\mathrm{II}}-u_{\mathrm{II}}^k,\ v_{\mathrm{I}}),$$
$$a(u_{\mathrm{II}}^{k+1}-u_{\mathrm{II}},v_{\mathrm{II}})=-a(u_{\mathrm{I}}^{k+1}-u_{\mathrm{I}},v_{\mathrm{II}})=a(u_{\mathrm{I}}-u_{\mathrm{I}}^{k+1},v_{\mathrm{II}})。$$

令 $v_{\mathrm{I}}=u_{\mathrm{I}}^{k+1}-u_{\mathrm{I}}$，$v_{\mathrm{II}}=u_{\mathrm{II}}^{k+1}-u_{\mathrm{II}}$ 且利用不等式(6.3)，有

$$\|u_{\mathrm{I}}^{k+1}-u_{\mathrm{I}}\|\leqslant\gamma\|u_{\mathrm{II}}^k-u_{\mathrm{II}}\|,$$
$$\|u_{\mathrm{II}}^{k+1}-u_{\mathrm{II}}\|\leqslant\gamma\|u_{\mathrm{I}}^{k+1}-u_{\mathrm{I}}\|,$$

故有

$$\|u_{\mathrm{I}}^{k+1}-u_{\mathrm{I}}\|\leqslant\gamma^2\|u_{\mathrm{I}}^k-u_{\mathrm{I}}\|,$$
$$\|u_{\mathrm{II}}^{k+1}-u_{\mathrm{II}}\|\leqslant\gamma^2\|u_{\mathrm{II}}^k-u_{\mathrm{II}}\|。\square$$

读者不难发现，由(6.6)式和(6.9)式所描述的分裂算法一和算法二实质上是求解有限元方程(6.4)的块 Jacobi 和块 GS 迭代方法，自然也会想到构造块 SOR 方法，所以有

定理6.3 若(6.3)式成立，$(\tilde{u}_{\mathrm{I}}^k,\tilde{u}_{\mathrm{II}}^k)^T$ 由(6.8)式或(6.9)式确定，则解有限元方程(6.2)或(6.4)的

分裂算法三：

$$\begin{cases} U_{\mathrm{I}}^{k+1} = (1-\omega)U_{\mathrm{I}}^{k} + \omega\widetilde{U}_{\mathrm{I}}^{k}, \\ U_{\mathrm{II}}^{k+1} = (1-\omega)U_{\mathrm{II}}^{k} + \omega\widetilde{U}_{\mathrm{II}}^{k}, \end{cases} \qquad (0<\omega\leqslant 2) \quad (6.10)$$

当 $\omega_{opt} = \dfrac{2}{1+\sqrt{1-\gamma^2}}$ 时，具有 h 无关的收敛因子$(1-\sqrt{1-r^2})/(1+\sqrt{1-\gamma^2})$，其中$\widetilde{U}_{\mathrm{I}}^{k}$和$\widetilde{U}_{\mathrm{II}}^{k}$由分裂算法二提供。

值得指出的是，只要 Ω_h 分划正规，V_h 设计合理，(6.3) 式就能成立且 γ 与 h 无关，从而分裂算法的敛速也与 h 无关。这是与经典迭代法的敛速与 h 有关，即与网格分划的细密程度有关的结论是截然不同的。另外，分裂算法将有限元方程分裂为两个低维的方程组，从而数值求解的计算量大为减少；分裂算法在 V_{I} 和 V_{II}上交错迭代，迭代过程可与节点编号无关(参见第五章§4)，这对减少计算存贮量也是极有益的。

§3 逐层分裂迭代法

为了更有效地发挥嵌套式迭代原则、分裂算法、外推技术、超收敛性技术和多层网格方法在数值求解有限元方程中的作用，本节介绍一种求解有限元方程的逐层分裂迭代法。

设用逐步加密法剖分区域 Ω (参见第五章§2)，就得到网格层序列$\{\Omega_i\}$，$i=1,2,\cdots,l$，在每一个网格层 Ω_i 上对应着一个与(6.2)式相应的有限元方程

$$a(u_i,\ v_i) = (f,\ v_i), \qquad \forall v_i \in V_h(\Omega_i) \qquad (6.11)$$

这里 $u_i = u_i(\Omega_i) = u_i(\Omega_{i-1}) \cup u_i(\Omega_i/\Omega_{i-1})$。

因为收缩数 γ 只与微分方程系数、网格分划规则和基函数的选取有关，这三个因素对于不同网格层可以保持一致，所以在这种情况下各个网格层 $\Omega_i(i=1,2,\cdots,l)$上的收缩数都是相同的。同样，若(6.3)式成立，则在每一网格层上都可以用分裂算法。此时，在(6.11)式中有

$$u_i(\Omega_i) = u_i(\Omega_{i-1}) \cup u_i(\Omega_i/\Omega_{i-1}),$$

$u_l(\Omega_{l-1}) \in V_{\mathrm{I}} = \mathrm{span}\{$单元顶节点基函数$\}$，

$u_l(\Omega_l/\Omega_{l-1}) \in V_{\mathrm{II}} = \mathrm{span}\{$单元边中点基函数$\}$。

在细网格层 Ω_l 上求解有限元方程(6.11)的逐层分裂迭代法具体步骤如下：

第一步　用一般迭代法或分裂迭代法解 $a(u_i, v_i) = (f, v_i)$，$i = 1, 2$，得 $u_1(\Omega_1)$和 $u_2(\Omega_2)$；

第二步　利用 $u_1(\Omega_1)$和 $u_2(\Omega_2)$ 外推，得$\tilde{u}_2(\Omega_2)$，且令 $\tilde{u}_3(\Omega_2) = \tilde{u}_2(\Omega_2)$；

第三步　利用 $\tilde{u}_3(\Omega_2)$内插，得 $\tilde{u}_3(\Omega_3/\Omega_2)$；

第四步　在 Ω_3/Ω_2和 Ω_2上用分裂算法一(或分裂算法二、三)作交错迭代；

第五步　在 $\Omega_i(i \geqslant 3)$层上重复第二、三、四步所述过程，直至求出 $u_l(\Omega_l)$；

第六步　最后在 Ω_l(利用 Ω_{l-1})层上使用外推和超收敛性技术，获得 Ω_l 层上高精度的有限元位移值和导数值。

实现该算法的计算框图见图6.1。

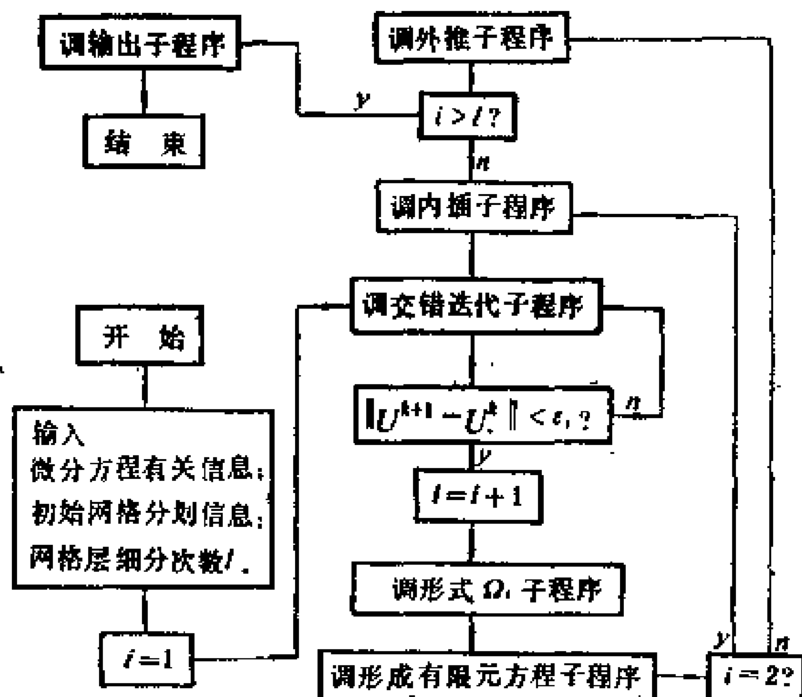


图　6.1

值得指出的是，该算法过程中的外推都是根据有限元解的渐近展开式(5.5)进行的，因此，对线性元和二次元要区别对待。另外，图 6.1 适用于一般三角形分划情形的逐层分裂迭代法，若是直角三角形分划，则图6.1中的“调形成 Ω_i 子程序”和“调形成有限元方程子程序”这两个模块可以省略。

关于逐层分裂迭代法的效果，有

定理6.4 若(6.3)式成立，则逐层分裂迭代法解 Ω_l 层上有限元方程(6.11)的敛速与 h 无关，总计算量为 $O(N_l)$。

证明 因为线性元或二次元的误差估计式 $\|u_i - u\| = O(h_i^{2\sim 3})$，所以有

$$\|u_{i-1} - u\| / \|u_i - u\| \approx 4 \sim 8, \quad i = 2, \cdots, l;$$

又由逐层分裂迭代法的第二、三步骤，有

$$\|u_i^0 - u\| \leqslant \|u_{i-1} - u\|, \quad i = 2, \cdots, l,$$

在极端的情况下，设在 Ω_i 层上经子空间分裂交错迭代 k 步后，数值解 u_i^k 已是(6.11)式的准确解(数值解已无法改进机器精度)，那么利用分裂迭代的 h 无关收敛性（收敛因子记为 ρ），则有

$$\|u_i - u\| = \|u_i^k - u\| = \rho^k \|u_i^0 - u\| \leqslant \rho^k \|u_{i-1} - u\|, \quad i = 2, \cdots, l.$$

所以，在任一网格层 $\Omega_i (2 \leqslant i \leqslant l)$ 上的交错迭代步数有保守估计：

$$k \leqslant K = \left| \lg \frac{1}{4} / \lg \rho \right|. \tag{6.12a}$$

再考虑到 Ω_i 层上施行内插和外推的工作量至多各相当于 Ω_i 层上的一个迭代步，故在 Ω_i 层上的全部计算量至多相当于该层上$(K+2)$步迭代的计算量。

因为 $N_{i-1}/N_i \approx 1/4 \sim 1/3 (i = 2, \cdots, l)$，所以在 $\Omega_1, \Omega_2, \cdots, \Omega_l$ 层上施行逐层分裂迭代法所需总计算量至多相当于 Ω_l 层上 $\tilde{K}$ 步迭代的工作量：

$$\tilde{K} \leqslant (K+2)\left(1 + \frac{1}{3} + \frac{1}{3^2} + \cdots\right) = \frac{3}{2}(K+2). \tag{6.12b}$$

最后考虑到 Ω_l 层上每迭代步的计算量为 $O(N_l)$，就可确定出使用逐层分裂迭代法的总计算量为 $O(N_l)$。□

下面具体估算用逐层分裂迭代法和单用分裂迭代法在细网层 Ω_l 上求解(6.11)式所需的总计算量，再把它折算成 Ω_l 层上迭代步数并列于表6.1中。表中所用到的线性元、二次元的设计及其相应的收缩数见§4和§5。

对于逐层分裂迭代法，我们用(6.12a)和(6.12b)式估计之。对于单用分裂迭代法求解的情形，我们假设 $h_l\approx10^{-3}$，$\|u-u_l^h\|\approx h_l^2\approx10^{-6}$，$\|u-u_l^0\|\approx1$，再估计出相应的迭代步数 k。这些列于表 6.1 中的数据已很清楚地表明了逐层分裂迭代法的高效率特征。逐层分裂迭代法的高精度特征是由细网层上外推和超收敛性技术（步骤六）保证的。值得指出的是，若用经典迭代法求解用通常方法得到的有限元方程，则花费的工作量大，其效果比单用分裂迭代法时的效果还要差很多。

表 6.1 逐层分裂法和单用分裂法工作量比较

数值方法 / 迭代次数 / 有限元设计	逐层分裂法			单用分裂法		
	分裂算法一	分裂算法二	分裂算法三	分裂算法一	分裂算法二	分裂算法三
一次元$\left(r=\frac{\sqrt{2}}{2}\right)$	12.0	7.5	4.8	53.2	26.6	10.5
二次元 I$\left(r=\frac{\sqrt{6}}{3}\right)$	16.8	9.9	5.4	90.6	45.0	14.0
二次元 Ⅲ$\left(r=\frac{\sqrt{8}}{8}\right)$	8.7	5.9	4.7	33.5	16.8	10.0

§4 适合强 Schwarz 不等式的三角线性元

强 Schwarz 不等式(6.3)是分裂算法和逐层分裂算法的核

心，本节将设计适合(6.3)式的线性有限元空间。

首先说明，强 Schwarz 不等式(6.3)的设计可以归结为某一个单元上的相应问题。

事实上，在任一网格层 Ω_l 上的有限元方程(6.11) 都是由所有单元 e 上的相应情形组合而成的，即

$$a(u,v)=\sum_e a_e(u,v),\quad (f,v)=\sum_e (f,v)_e.$$

同样，与(6.11)式相应的矩阵表示式 (6.4)，也是所有单元 e 上的子刚阵和子右端合成的结果（参见第二篇§2.4），即

$$\begin{pmatrix}A_h & C_h\\ C_h^T & B_h\end{pmatrix}\begin{pmatrix}U_{\rm I}\\ U_{\rm II}\end{pmatrix}=\sum_e\begin{pmatrix}A_h^e & C_h^e\\ C_h^{eT} & B_h^e\end{pmatrix}\begin{pmatrix}U_{\rm I}^e\\ U_{\rm II}^e\end{pmatrix}=\sum_e\begin{pmatrix}F_{\rm I}^e\\ F_{\rm II}^e\end{pmatrix}=\begin{pmatrix}F_{\rm I}\\ F_{\rm II}\end{pmatrix},$$

上面两式中，带有角标 e 的矩阵和向量都表示限制在单元 e 上的情形。这样就有

$$\begin{aligned}|a(u,v)|=\Big|\sum_e a_e(u,v)\Big|&\leqslant\sum_e \gamma_e a_e^{1/2}(u,u)a_e^{1/2}(v,v)\\ &\leqslant(\max_e(\gamma_e))\cdot a^{1/2}(u,u)a^{1/2}(v,v),\end{aligned}$$

其中

$$\begin{cases}\gamma_e=\sup\{|a_e(u,v)|\,\big|\,u\in V_{\rm I}^e,\ v\in V_{\rm II}^e,\ a_e(u,u)=a_e(v,v)=1\}\\ \gamma\leqslant\max\limits_e\gamma_e.\end{cases}\tag{6.12c}$$

因此，利用 (6.12c) 式不仅可验证有限元空间的设计是否符合(6.3)式，而且可以具体计算出收缩数 γ_e 和 γ。

还可以用代数办法求得收缩数 γ。例如，若设 A_h 和 B_h 非奇异，则由定理6.2知分裂算法二：

$$\begin{cases}A_h U_{\rm I}^{k+1}=F_{\rm I}-C_h U_{\rm II}^k,\\ B_h U_{\rm II}^{k+1}=F_{\rm II}-C_h^T U_{\rm I}^{k+1},\end{cases}$$

具有收敛因子 γ^2，从而有

$$U_{\rm I}^{k+1}-U_{\rm I}=-A_h^{-1}C_h(U_{\rm II}^k-U_{\rm II})=A_h^{-1}C_hB_h^{-1}C_h^T(U_{\rm I}^k-U_{\rm I}),$$
$$U_{\rm II}^{k+1}-U_{\rm II}=-B_h^{-1}C_h^T(U_{\rm I}^{k+1}-U_{\rm I})=B_h^{-1}C_h^TA_h^{-1}C_h(U_{\rm II}^k-U_{\rm II}),$$

故有

$$\gamma^2=\rho(A_h^{-1}C_hB_h^{-1}C_h^T)=\rho(B_h^{-1}C_h^TA_h^{-1}C_h). \qquad (6.13a)$$

若 A_h 奇异，B_h 非奇异(即 $a(\cdot,\cdot)$在 V_{II}中正定，在 V_{I}中半正定)，则γ^2可通过特征值问题

$$C_hB_h^{-1}C_h^Tx=\lambda A_hx \qquad (6.13b)$$

求得。当然，人们可利用(6.13a)和(6.13b)式在子刚阵中的表现先求出γ_e，然后再求得收缩数γ。

值得指出的是，对于一致网格分划情形，在区域内部单元上求得的收缩数$\gamma_e\geqslant\gamma$。若令$\gamma_e=\gamma$且用其来估计分裂算法的收敛因子，那么这相当于求解 Neumann 边值问题的情形；对于强制边值条件的情形，一般$\gamma_e>\gamma$；对混合边值条件的情形，一般有$\gamma_e\geqslant\gamma$。

下面构造适合强 Schwarz 不等式的线性元，并用(6.13a)和(6.13b)式计算出收缩数γ_e。

在任一单元 $e(\triangle a_1a_2a_3)$上，参见图6.2，设单元顶点线性基函数记为 φ_i，单元边中点线性基函数记为 ψ_j，且

$$\varphi_i(x,y)=L_i(x,y),\ (x,y)\in e,\ i=1,2,3,$$

$$\psi_j(x,y)=L_j(x,y),\ (x,y)\in e_m,m=1,\cdots,4,\ j=1,2,3.$$

即，φ_i是定义在单元 e 上的面积坐标，ψ_j 是定义在子单元 e_m 上的面积坐标。显然，V_h^e 是线性元空间，且

$$V_h^e=V_{\text{I}}^e\oplus V_{\text{II}}^e,$$

$$V_{\text{I}}^e=\text{span}\{\varphi_i\mid i=1,2,3\},$$

$$V_{\text{II}}^e=\text{span}\{\psi_j\mid j=1,2,3\}.$$

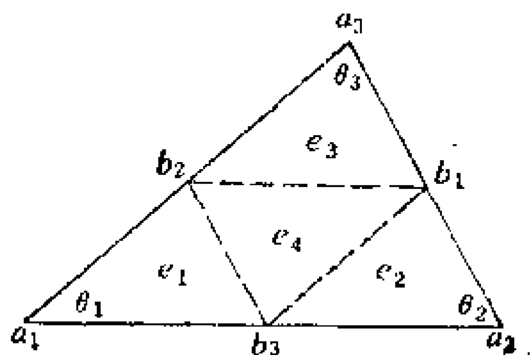


图6.2

设单元 e 的面积为 S，e_i的面积为 $S/4$，相应于

$$a_e(u,v)=\int_e\nabla u\nabla v\,dxdy+q\int_e uv\,dxdy$$

的单元刚度矩阵 $\begin{bmatrix} A & C \\ C^T & B \end{bmatrix}$ 可相应写为

$$A = A_1 + tA_2,\ B = B_1 + tB_2,\ C = C_1 + tC_2,\ t = qS,$$

还记 θ_i 是单元 e 在顶点 a_i 处的三个角度参数，

$$c_i = \operatorname{ctg}\theta_i,\quad c = \sum_{i=1}^{3} c_i,\quad d = \sum_{i=1}^{3} \cos^2\theta_i,\quad i = 1,\ 2,\ 3.$$

有

$$A_1 = \frac{1}{2}\begin{pmatrix} c_2 + c_3 & -c_3 & -c_2 \\ -c_3 & c_3 + c_1 & -c_1 \\ -c_2 & -c_1 & c_1 + c_2 \end{pmatrix},\quad B_1 = \begin{pmatrix} c & -c_3 & -c_2 \\ -c_3 & c & -c_1 \\ -c_2 & -c_1 & c \end{pmatrix},$$

$$C_1 = -A_1; \tag{6.14}$$

$$A_2 = \frac{1}{12}\begin{pmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{pmatrix},\quad B_2 = \frac{1}{24}\begin{pmatrix} 3 & 1 & 1 \\ 1 & 3 & 1 \\ 1 & 1 & 3 \end{pmatrix},\quad C_2 = \frac{1}{48}\begin{pmatrix} 2 & 5 & 5 \\ 5 & 2 & 5 \\ 5 & 5 & 2 \end{pmatrix}. \tag{6.15}$$

下面验证(6.14)和(6.15)式的正确性。因为

$$A_1 = \left(\int_e \nabla\varphi_i \nabla\varphi_j dxdy\right)_{3\times 3},\qquad B_1 = \left(\int_e \nabla\psi_i \nabla\psi_j dxdy\right)_{3\times 3},$$

$$C_1 = \left(\int_e \nabla\varphi_i \nabla\psi_j dxdy\right)_{3\times 3},\qquad A_2 = \left(\frac{1}{S}\int_e \varphi_i\varphi_j dxdy\right)_{3\times 3},$$

$$B_2 = \left(\frac{1}{S}\int_e \psi_i\psi_j dxdy\right)_{3\times 3},\qquad C_2 = \left(\frac{1}{S}\int_e \varphi_i\psi_j dxdy\right)_{3\times 3},$$

$$i,\ j = 1,\ 2,\ 3.$$

所以，由第二篇公式(2.31)，就可直接计算出

$$\int_e \nabla\varphi_1 \nabla\varphi_1 dxdy = |a_2a_3|^2/4S = \frac{1}{2}(c_2 + c_3),$$

$$\int_e \nabla\varphi_1 \nabla\varphi_2 dxdy = \overrightarrow{a_2a_3} \cdot \overrightarrow{a_3a_1}/4S = -c_3/2,$$

$$\int_e \nabla\psi_1\nabla\psi_1 dxdy = \frac{|a_3b_2|^2}{S} + \frac{|b_2b_3|^2}{S} + \frac{|b_3a_2|^2}{S} = c,$$

$$\int_e \nabla\psi_1\nabla\psi_2 dxdy = \frac{\overrightarrow{b_2a_3}\cdot\overrightarrow{a_3b_1}}{S} + \frac{\overrightarrow{b_2b_3}\cdot\overrightarrow{b_3b_1}}{S} = -c_3,$$

$$\int_e \nabla\varphi_1\nabla\psi_2 dxdy = \int_e \nabla\left(\psi_{a_1} + \frac{1}{2}\psi_2 + \frac{1}{2}\psi_3\right)\nabla\psi_2 dxdy = \frac{1}{2}c_3,$$

$$\int_e \varphi_2\psi_3 dx\,dy = \int_e\left(\psi_{a_2} + \frac{1}{2}\psi_1 + \frac{1}{2}\psi_3\right)\psi_3 dxdy = \frac{5}{48}S.$$

同样，也可以计算验证(6.14)和(6.15)式中其它元素的正确性。

下面计算关于$\int_e \nabla u\nabla v dxdy$的收缩数$\gamma_e$。

由(6.13a)和(6.13b)知，求γ_e^2就是要解特征值问题：

$A_1B_1^{-1}A_1x = \lambda A_1x$ 或 $B_1^{-1}A_1y = \lambda y$。

若记 $D_c = \mathrm{diag}(c_1, c_2, c_3)$ 且利用关系 $A_1 = \frac{1}{2}B_1 - \frac{1}{2}D_c$，则上式可写为

$$\frac{1}{2}(I - B^{-1}D_c)y = \lambda y,$$

所以

$$\gamma_e^2 = \frac{1}{2}(1 - \lambda_{\min}(B^{-1}D_c)). \tag{6.16}$$

利用三角恒等式

$$\frac{c_1c_2c_3}{c} = \frac{1-d}{3-d},$$

并经计算可求得适合$|D_c - \lambda B_1| = 0$的

$$\lambda_1 = 1,\quad \lambda_2 = \frac{1}{4} + \sqrt{4d-3},\quad \lambda_3 = \frac{1}{4} - \sqrt{4d-3}.$$

将λ_3代入(6.16)式，就得到收缩数

$$\gamma_e^2 = \frac{3}{8} + \frac{1}{4}\sqrt{d - \frac{3}{4}}. \tag{6.17}$$

现将γ_e和d的关系曲线绘置于图6.3中。

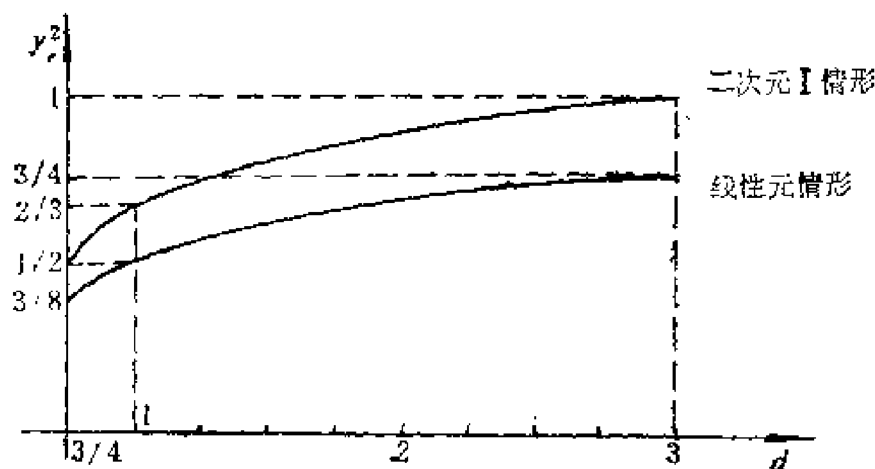


图 6.3

下面计算关于$\int_e uvdxdy$的收缩数。

由(6.15)式知，A_2、B_2和C_2三个矩阵都是形如

$$\alpha I+\beta EE^T,\qquad E^T=(1,\ 1,\ 1)$$

的矩阵，因此只要求出EE^T的特征值，即可利用有关定理求得$\begin{pmatrix}A_2 & C_2\\ C_2^T & B_2\end{pmatrix}$的特征值$\lambda$，显然

若 $A_2=\alpha_1 I+\beta_1 EE^T, B_2=\alpha_2 I+\beta_2 EE^T, C_2=\alpha_3 I+\beta_3 EE^T$,

则 $\lambda_1=\dfrac{(\alpha_3+3\beta_3)^2}{(\alpha_1+3\beta_1)(\alpha_2+3\beta_2)}$, $\lambda_2=\dfrac{\alpha_3^2}{\alpha_1\alpha_2}$,

$$\gamma_e^2=\max\{\lambda_1,\ \lambda_2\},\tag{6.18}$$

代入(6.15)式的A_2、B_2和C_2的有关参数，就得到

$$\gamma_e^2=\max\left\{\frac{9}{16},\ \frac{9}{16}\right\}=\frac{9}{16}.\tag{6.19}$$

下面计算正三角形单元上的关于$\int_e \nabla u\nabla vdxdy+q\int_e uvdxdy$的收缩数。

对于三角单元的任意角度参数情形，不便描述γ_e与某个参

数的依赖关系。对于正三角形单元情形，A_1、B_1、C_1、A_2、B_2和C_2都是形如$\bar{\alpha}I+\bar{\beta}EE^T$的矩阵，此时容易寻找出$\gamma_e$与$t$的依赖关系。

由(6.14)和(6.15)式，有

$$A=\bar{\alpha}_1 I+\bar{\beta}_1 EE^T,\quad B=\bar{\alpha}_2+\bar{\beta}_2 EE^T,\quad C=\bar{\alpha}_3+\bar{\beta}_3 EE^T$$

$$\bar{\alpha}_1=\frac{\sqrt{3}}{2}+\frac{t}{12},\qquad \bar{\alpha}_2=\frac{4}{3}\sqrt{3}+\frac{t}{12},\qquad \bar{\alpha}_3=-\left(\frac{\sqrt{3}}{2}+\frac{t}{16}\right),$$

$$\bar{\beta}_1=-\frac{\sqrt{3}}{6}+\frac{t}{12},\qquad \bar{\beta}_2=-\frac{\sqrt{3}}{3}+\frac{t}{24},\qquad \bar{\beta}_3=\frac{\sqrt{3}}{6}+\frac{5}{48}t,$$

再利用(6.18)式，就得到

$$\begin{cases}\gamma_e^2=\max\{\lambda_1,\ \lambda_2\},\\ \lambda_1=\dfrac{9}{10}\,\dfrac{t}{t+\dfrac{8}{15}\sqrt{3}},\quad \lambda_2=\dfrac{9}{16}\,\dfrac{\left(t+\dfrac{8}{3}\sqrt{3}\right)^2}{(t+2\sqrt{3})\left(t+\dfrac{16}{3}\sqrt{3}\right)}.\end{cases} \tag{6.20}$$

我们将γ_e^2与$qh^2$$\left(\text{正三角形边长为}h,\ S=\dfrac{\sqrt{3}}{2}h^2, qh^2=\dfrac{2\sqrt{3}}{3}t\right)$的关系曲线绘置于图6.4中。

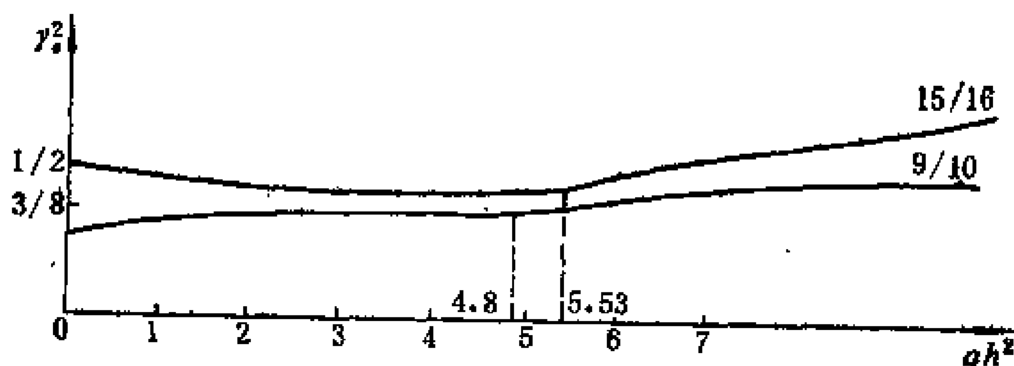


图 6.4

当然，图6.4中的曲线只是表明了正三角形单元这种特殊情

形下的收缩数，对于任意角度参数关于 $p(x,y)\int_e \nabla u \nabla v dxdy + q(x,y)\int_e uvdxdy$ 的收缩数，不仿借助于数值计算描绘出近似曲线，这对设计和分析分裂算法及逐层分裂迭代法将是有指导意义的。

§5 适合强 Schwarz 不等式的三角二次元

本节介绍三种适合强 Schwarz 不等式的二次元，并计算其相应的收缩数。

二次元 I

如图6.2所示，设

顶节点基函数，$\varphi_i = L_i$，$i=1, 2, 3$；

边中点基函数，$\psi_1 = 4L_2L_3$，$\psi_2 = 4L_1L_3$，$\psi_3 = 4L_1L_2$。其中 $L_i(i=1,2,3)$ 是定义在单元 e 上的面积坐标，则 $V_h^e = V_{\rm I}^e \oplus V_{\rm II}^e$，$V_{\rm I}^e = {\rm span}\{\varphi_i | i=1,2,3\}$，$V_{\rm II}^e = {\rm span}\{\psi_i | i=1, 2, 3\}$。

如同§4的(6.14)和(6.15)式，相应于二次元 I，有

$$A_1 = \frac{1}{2}\begin{pmatrix} c_2+c_3 & -c_3 & -c_2 \\ -c_3 & c_3+c_1 & -c_1 \\ -c_2 & -c_1 & c_1+c_2 \end{pmatrix},\quad B_1 = \frac{4}{3}\begin{pmatrix} c & -c_3 & -c_2 \\ -c_3 & c & -c_1 \\ -c_2 & -c_1 & c \end{pmatrix},$$

$$C_1 = \frac{-4}{3}A_1; \tag{6.21}$$

$$A_2 = \frac{1}{12}\begin{pmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{pmatrix},\quad B_2 = \frac{8}{90}\begin{pmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{pmatrix},\quad C_2 = \frac{1}{15}\begin{pmatrix} 1 & 2 & 2 \\ 2 & 1 & 2 \\ 2 & 2 & 1 \end{pmatrix}. \tag{6.22}$$

仍按照§4中的办法，利用关系 $A_1 = \frac{3}{8}B_1 - \frac{1}{2}D_c$，通过求解特征问题

$$\frac{16}{9}B_1^{-1}A_1 y = \frac{16}{9}B_1^{-1}\left(\frac{3}{8}B_1 - \frac{1}{2}D_c\right) = \lambda y,$$

即有

$$\gamma_e^2=\frac{2}{3}\left(1-\lambda_{\min}\left(\frac{4}{3}B_1^{-1}D_e\right)\right)$$

其中

$$\lambda_1=1,\quad \lambda_2=\frac{1}{4}+\frac{1}{4}\sqrt{4d-3},\quad \lambda_3=\frac{1}{4}-\frac{1}{4}\sqrt{4d-3},$$

这样可求得二次 I 关于$\int_e \nabla u\nabla v dxdy$ 的收缩数

$$\gamma_e^2=\frac{1}{2}+\frac{1}{3}\sqrt{d-\frac{3}{4}}, \tag{6.23}$$

还将 γ_e^2随 d 的变化曲线绘置于图6.3中。

按§4中的分析办法，利用(6.18)式和(6.22)式，也可计算出二次元 I 关于$\int_e uvdxdy$ 的收缩数

$$\gamma_e^2=\max\left\{\frac{15}{16},\ \frac{3}{5}\right\}=\frac{15}{16}. \tag{6.24}$$

最后就正三角单元情形，把二次元 I 关于$\int_e \nabla u\nabla v dxdy+q\int_e uvdxdy$的收缩数

$$\begin{cases}\gamma_e^2=\max\{\lambda_1,\ \lambda_2\},\\ \lambda_1=\dfrac{15}{16}\dfrac{t}{t+\dfrac{5}{4}\sqrt{3}},\quad \lambda_2=\dfrac{3}{5}\dfrac{(t+10\sqrt{3})^2}{(t+6\sqrt{3})(t+20\sqrt{3})}\end{cases} \tag{6.25}$$

的变化曲线绘置于图 6.4 中。

二次元Ⅱ

如图 6.5 所示，设三角单元 e 的面积为 S，$|A_2A_3|=a$，$|A_3A_1|=b$，$|A_1A_2|=c$，$|AD|=m$，$|A_2D|=t_1$，$|Dt_3|=t_2$。

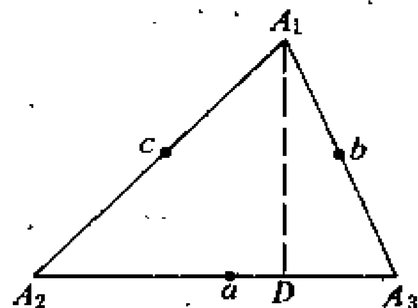


图 6.5

在单元 e 上用下面节点基函数构造二次元：

顶节点基函数 $\varphi_1=1$，$\varphi_2=x$，$\varphi_3=y$ 张成子空间 V_{I}；

边中点基函数 $\psi_1=4L_2L_3$，$\psi_2=4L_1L_3$，$\psi_3=4L_1L_2$ 张成子空间 V_{II}。

对 Poisson 方程，可计算子刚阵各元素

$$
\begin{cases}
a_e(\varphi_1,\varphi_1)=a_e(\varphi_1,\varphi_2)=a_e(\varphi_1,\varphi_3) \\
\qquad =a_e(\varphi_1,\psi_1)=a_e(\varphi_1,\psi_2)=a_e(\varphi_1,\psi_3)=0, \\
a_e(\varphi_2,\varphi_2)=S=a_e(\varphi_3,\varphi_3), \\
a_e(\varphi_2,\varphi_3)=a_e(\varphi_2,\psi_1)=0, \\
a_e(\varphi_2,\psi_2)=a_e(\varphi_3,\psi_3)=\dfrac{4}{3}mS, \\
a_e(\psi_1,\psi_1)=\dfrac{8}{3}[b^2+c^2-(m^2-t_1t_2)]S, \\
a_e(\psi_1,\psi_2)=-\dfrac{8}{3}at_2S, \\
a_e(\psi_1,\psi_3)=-\dfrac{8}{3}at_1S, \\
a_e(\psi_2,\psi_2)=\dfrac{8}{3}[a^2+c^2-t_1a]S, \\
a_e(\psi_2,\psi_3)=-\dfrac{8}{3}[m^2-t_1t_2]S, \\
a_e(\psi_3,\psi_3)=\dfrac{4}{3}t_1S.
\end{cases}
\tag{6.26}
$$

设
$$
u=\sum_{i=1}^{3}\alpha_i\varphi_i\in V_{\text{I}},\quad v=\sum_{i=1}^{3}\beta_i\psi_i\in V_{\text{II}},
\tag{6.27}
$$

则有
$$
\begin{cases}
a_e(u,u)=a_e\left(\sum\limits_{i=1}^{3}\alpha_i\varphi_i,\ \sum\limits_{i=1}^{3}\alpha_i\varphi_i\right), \\
a_e(v,v)=a_e\left(\sum\limits_{i=1}^{3}\beta_i\psi_i,\ \sum\limits_{i=1}^{3}\beta_i\psi_i\right), \\
a_e(u,v)=a_e\left(\sum\limits_{i=1}^{3}\alpha_i\varphi_i,\ \sum\limits_{i=1}^{3}\beta_i\psi_i\right).
\end{cases}
\tag{6.28}
$$

利用(6.26)～(6.28)式和 Lagrange 乘子法（按定义(6.12c)式）即可算出 γ_e。令

$$R = |a_e(u,v)| + \lambda_1[a_e(u,u) - 1] + \lambda_2[a_e(v,v) - 1], \qquad (6.29)$$

则由

$$\begin{cases} R'_{\alpha_i} = R'_{\beta_i} = 0, \quad i = 1, \ 2, \ 3, \\ a_e(u,u) = a_e(v,v) = 1, \end{cases} \qquad (6.30)$$

可计算出 α_i 和 $\beta_i(i = 1, \ 2, \ 3)$以及 λ_1和 λ_2，再将这 8 个参数代入(6.29)式，即可计算出

$$\gamma_e = \max R. \qquad (6.31)$$

当然，人们也可以利用各种最优化方法数值求解(6.12)式，也可求得γ_e。

例如，若 e 是等腰直角三角单元，即 $c = a = m = t_2 = 1, t_1 = 0$，则由上面诸式可算得 $\gamma_e = \sqrt{2/3}$。

若 e 是正三角形单元，即 $a = b = c = 1$，$t_1 = t_2 = 1/2$，$m = \sqrt{3}/2$，则由(6.27)～(6.31)诸式有

$$\begin{aligned} R = & \frac{2\sqrt{3}}{3}\alpha_2\beta_2 - \frac{2\sqrt{3}}{3}\alpha_2\beta_3 - \frac{4}{3}\alpha_3\beta_1 + \frac{2}{3}\alpha_3\beta_3 + \frac{2}{3}\alpha_3\beta_2 \\ & + \lambda_1(\alpha_2^2 + \alpha_3^2) + \lambda_2\Big[4(\beta_1^2 + \beta_2^2 + \beta_3^2) \\ & - \frac{8}{3}(\beta_1\beta_2 + \beta_2\beta_3 + \beta_3\beta_1)\Big], \end{aligned}$$

$$R'_{\alpha_2} = 0, \text{ 即 } \quad \frac{2\sqrt{3}}{3}\beta_2 - \frac{2\sqrt{3}}{3}\beta_3 + 2\lambda_1\alpha_2 = 0,$$

$$R'_{\alpha_3} = 0, \text{ 即 } \quad -\frac{4}{3}\beta_1 + \frac{2}{3}\beta_2 + \frac{2}{3}\beta_3 + 2\lambda_1\alpha_3 = 0,$$

$$R'_{\beta_1} = 0, \text{ 即 } \quad -\frac{4}{3}\alpha_3 + \lambda_2\Big[8\beta_1 - \frac{8}{3}\beta_2 - \frac{8}{3}\beta_3\Big] = 0,$$

$$R'_{\beta_2} = 0, \text{ 即 } \quad \frac{2\sqrt{3}}{3}\alpha_2 + \frac{2}{3}\alpha_3 + \lambda_2\Big[8\beta_2 - \frac{8}{3}\beta_1 - \frac{8}{3}\beta_3\Big] = 0,$$

$$R'_{\beta_3} = 0, \text{ 即 } \quad -\frac{2\sqrt{3}}{3}\alpha_2 + \frac{2}{3}\alpha_3 + \lambda_2\Big[8\beta_3 - \frac{8}{3}\beta_1 - \frac{8}{3}\beta_2\Big] = 0.$$

由后三个方程相加，且因 $\lambda_2 \neq 0$，就有

$$\beta_1 + \beta_2 + \beta_3 = 0,$$

即

$$\beta_1^2+\beta_2^2+\beta_3^2=-2\beta_1\beta_2-2\beta_2\beta_3-2\beta_3\beta_1.$$

最后，利用上面两式，有

$$|a_e(u,v)|^2=\left[\frac{2\sqrt{3}}{3}\alpha_2\beta_2-\frac{2\sqrt{3}}{3}\alpha_2\beta_3-\frac{4}{3}\alpha_3\beta_1-\frac{2}{3}\alpha_3\beta_1\right]^2$$

$$=4\alpha_3^2\beta_1^2+\frac{8}{3}\sqrt{3}\,(\alpha_2\beta_3\alpha_3\beta_1-\alpha_2\beta_2\alpha_3\beta_1)$$

$$+\frac{4}{3}(\alpha_2\beta_2-\alpha_2\beta_3)^2$$

$$\leqslant 4\alpha_3^2\beta_1^2+\frac{4}{3}\alpha_2^2\beta_1^2+4\alpha_3^2\beta_2^2+\frac{4}{3}\alpha_2^2\beta_1^2+4\alpha_3^2\beta_3^2$$

$$+\frac{8}{3}\alpha_2^2\beta_2^2+\frac{8}{3}\alpha_2^2\beta_3^2$$

$$\leqslant 4(\alpha_3^2\beta_1^2+\alpha_2^2\beta_3^2+\alpha_2^2\beta_1^2+\alpha_3^2\beta_3^2+\alpha_2^2\beta_2^2+\alpha_3^2\beta_2^2)$$

$$=4(\alpha_2^2+\alpha_3^2)(\beta_1^2+\beta_2^2+\beta_3^2)=\frac{3}{4}a_e(u,u)a_e(v,v).$$

所以，估算出二次元Ⅱ求解 **Poisson** 方程的收缩数 $\gamma_e\leqslant\sqrt{3}/2$.

二次元Ⅲ

在单元 e 上用下面节点基函数构造二次元：

顶节点基函数，$\varphi_1=2L_1^2-L_1$，$\varphi_2=2L_2^2-L_2$，$\varphi_3=2L_3^2-L_3$；

边中点基函数，$\psi_1=1$，$\psi_2=x$，$\psi_3=y$.

用(6.27)～(6.30)式，可计算出二次元Ⅲ（直角等腰三角形单元情形）求解 **Poisson** 方程的收缩数 $\gamma_e=\sqrt{3}/3$.

§6 RB 分划下线性元的收缩数

设 **RB** 分划下网格层序列按从粗到细的顺序排列为 Ω_1，Ω_2，…，Ω_l。此时，网格层仍有直和分解结构：$\Omega_1\cup(\Omega_2/\Omega_1)\cup(\Omega_3/\Omega_2)\cup\cdots\cup(\Omega_l/\Omega_{l-1})=\Omega_l$.

现就 $\Omega_h(=\Omega_1)$和 $\Omega_{\sqrt{2}h/2}(=\Omega_2)$两网层关系考虑之，其它任两个相邻网格层之关系完全相同，参见图6.6，点1、3、5、7、9、

11、13、15属于 Ω_1，偶数编号点属于 Ω_2/Ω_1，所以，$\Omega_2=\Omega_1\cup(\Omega_2/\Omega_1)$。

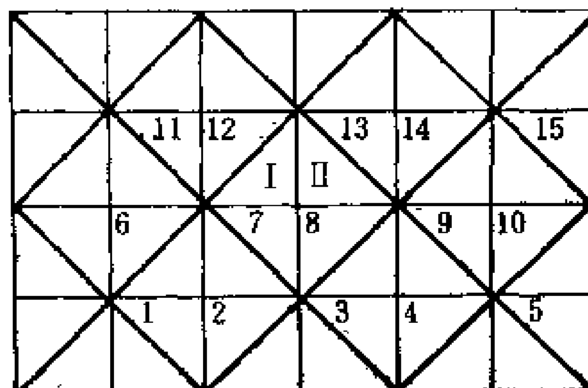


图 6.6

对于 Ω_1 上的点，采用 Ω_1 上的三角线性元节点基函数 φ_i；对于 Ω_2/Ω_1 上的那些点，采用 Ω_2 上的三角线性元节点基函数 ψ_i；相应地记 $u\in V_h(\Omega_1)$，$v\in V_h(\Omega_2/\Omega_1)$。

现限制在任一单元 $e\in\Omega_1$（不妨设 $\triangle_{7,9,13}$）上计算 $a_e(u,v)$ 的值。

因为 v 是分片线性函数且 $v(P_7)=v(P_9)=v(P_{13})=0$，所以有

$$u_\xi(\text{Ⅰ})=u_\xi(\text{Ⅱ}),\quad u_\eta(\text{Ⅰ})=u_\eta(\text{Ⅱ}),$$

$$v_\xi(\text{Ⅰ})=-v_\xi(\text{Ⅱ}),\quad v_\eta(\text{Ⅰ})=v_\eta(\text{Ⅱ}),\quad |v_\xi|=|v_\eta|,$$

从而有

$$\begin{aligned}|a_e(u,v)|&=\left|\iint_e(u_\xi v_\xi+u_\eta v_\eta)d\xi d\eta\right|\\&=\left|\iint_e u_\eta v_\eta d\xi d\eta\right|=\left(\int_e|u_\eta|^2d\xi d\eta\cdot\int_e|v_\eta|^2d\xi d\eta\right)^{1/2}\\&=\left(\int_e|u_\eta|^2d\xi\, d\eta\cdot\int_e\frac{1}{2}(|v_\xi|^2+|v_\eta|^2)d\xi d\eta\right)^{1/2}\\&\leqslant\frac{\sqrt{2}}{2}\left(\int_e(|u_\xi|^2+|u_\eta|^2)d\xi d\eta\int_e(|v_\xi|^2\right.\\&\qquad\left.+|v_\eta|^2)d\xi d\eta\right)^{1/2}.\end{aligned}$$

故 RB 分划下的三角线性元解 Poisson 方程的收缩数为 $\gamma_{*}=\sqrt{2}/2$。

作为本章结束，我们强调如下几个问题。

第一，前面几节只讨论了有限元的收缩数，没有介绍施行分裂算法的计算公式，即局部离散格式，读者在充分理解单元刚度矩阵和局部离散格式的区别与联系的基础上，可以利用单元刚度矩阵自行推导出在分裂算法中与各类不同点所对应的局部离散格式。

第二，研究矩形元的收缩数及其在分裂算法和逐层分裂算法中的应用也是很有意义的。研究办法仍同于三角元情形，即先作特殊的矩形元设计，写成相应的单元刚度矩阵，再用最优化方法或代数方法计算出矩形元的收缩数，最后利用单元刚度矩阵或不同类型点的局部离散格式，就可方便地进行分裂算法的数值计算。

第三，根据有限元格式和差分格式的某些统一性可以推知，对某些差分格式采用分裂算法应是可行的，有兴趣的读者可自行深入讨论之。

第四，对本章的特殊有限元构造，例如 §4 中的三角线性元等，按 §2 中的分裂算法计算，请读者注意，这样算出的 u_{I} 不是单元边中点处的有限元值，而是相对于 u_{I} 在该点处数值的校正量。

第七章　余量校正迭代方法

多层网格方法是求解离散方程$L_h u_h = f_h$的高效率方法，但用此方法并不能改善数值解的精度阶。使用外推技术可有效地提高此离散方程数值解的精度，但人们必须知道数值解u_h所具有的渐近展开式。另一种提高此离散方程数值解精度的办法是在进一步利用残差量校正技术的基础上形成的，称为**余量校正迭代(Iteration Defect Correct)方法**，简称为**IDC方法**。该方法最先由L. Fox, K. Böhmer, H. J. Stetter[41]~[44]等人于70年代初期开始总结研究的，现已成为数值求解线性和非线性问题的有效数值方法之一。

本章介绍IDC方法的原理、构造及其数值分析结果，还将介绍该方法与其它数值方法结合的新算法。

§1　余量校正迭代方法

设有线性椭圆型偏微分方程

$$Lu = f \tag{7.1}$$

的边值问题，与(7.1)式相应的离散方程是

$$L_h u_h = f_h \tag{7.2}$$

其中L_h的相容逼近阶是κ，求解(7.2)式容易且数值稳定。

正如第三章§1所述，经典迭代方法数值求解(7.2)式用到了余量校正原则。例如对任给的近似值u_h^l，经典迭代方法可表示为

$$\begin{cases} u_h^{l+1} = u_h^l + v_h^l, \\ Q_h v_h^l = d_h^l := f_h - L_h u_h^l, \end{cases}$$

其中Q_h是相对于L_h的简化离散算子，v_h^l称为**余量或校正量**。

收敛的迭代结果是 $u_h^i \to u_h$，$\|u-u_h\|=O(h^\kappa)$。

正如第四章所述，MG方法保持了一般迭代方法的余量校正原则，但把求解残差方程的工作转化到更粗的网格层上进行，从而提高了求解效率，然而MG方法并不能改进(7.2) 式数值解 u_h 的精度阶。

当然，(7.2)式数值解精度是由离散算子L_h的相容逼近阶决定的，为了获得更高精度的数值解，人们自然期望构造出更高相容阶的离散方程

$$\tilde{L}_h\tilde{u}_h=\tilde{f}_h, \tag{7.3}$$

其中 $\tilde{L}_h$ 的相容阶为 $\tilde{\kappa}(\tilde{\kappa}>\kappa)$。然而，相容阶越高，$\tilde{L}_h$ 构造越复杂，这不仅使数值求解 (7.3) 式的计算量增大，而且常常出现数值不稳定，甚至出现用通常方法不能求解的缺点。

IDC 方法的目的在于，充分利用数值解精度低但数值稳定的(7.2)式，且充分利用余量校正原则，顺利完成(7.3)式的数值求解工作。下面详细介绍 IDC 方法的原理。

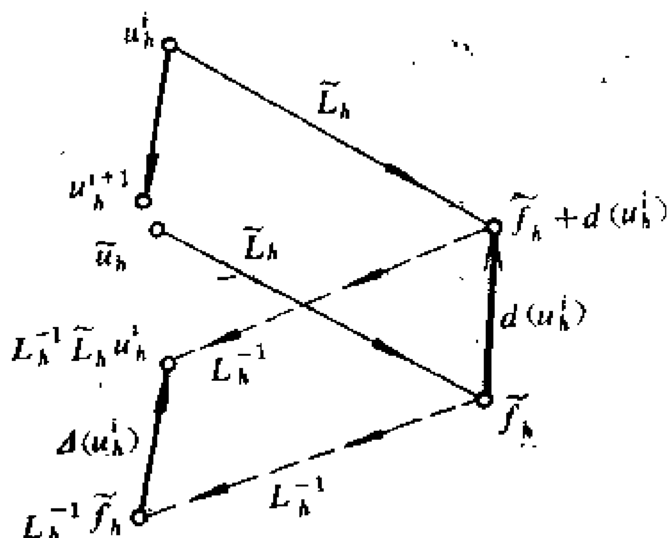


图 7.1

第一种办法：参见图 7.1，对任给的近似值 u_h^i，$\tilde{L}_hu_h^i$ 可看作是 (7.3) 式右端项 $\tilde{f}_h$ 的近似值，这时(7.3)式的右端项发生了

改变量

$$d(u_h^i)=\tilde{L}_h u_h^i-\tilde{f}_h.$$

再把 $d(u_h^i)$ 近似地看作是(7.2) 式右端项的改变量，从而可以求得与之相应的近似解改变量

$$\Delta(u_h^i)=L_h^{-1}d(u_h^i).$$

最后利用 $\Delta(u_h^i)$ 校正近似值 u_h^i 得新的近似值

$$\begin{aligned}u_h^{i+1}&=u_h^i-\Delta(u_h^i)\\&=u_h^i-[L_h^{-1}(\tilde{f}_h+d(u_h^i))-L_h^{-1}\tilde{f}_h]\\&=u_h^i-L_h^{-1}\tilde{L}_h u_h^i+L_h^{-1}\tilde{f}_h.\end{aligned}\tag{7.4}$$

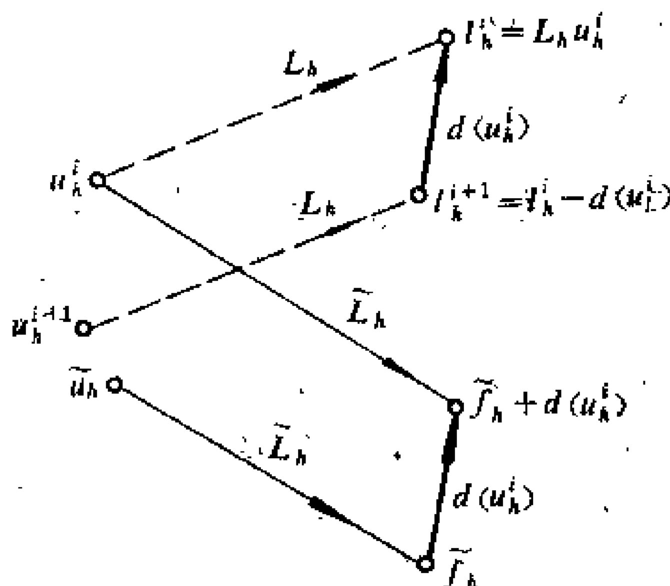


图 7.2

第二种办法：参见图 7.2。我们把近似解的改善过程看作是不断校正离散方程

$$L_h u_h^i=l_h^i$$

的右端项的过程。对任给的近似值 u_h^i，一方面可求出(7.2)式右端项与之相应的近似值 $L_h u_h^i$；另一方面，也可求出(7.3)式右端项的改变量 $d(u_h^i)$。先用 $d(u_h^i)$ 校正(7.2)式的右端项的近似值，有

$$l_h^{i+1}=l_h^i-d(u_h^i);$$

再求改变后的右端项对应的新的近似解

$$u_h^{i+1}=L_h^{-1}l_h^{i+1}.$$

所以我们有

$$\begin{aligned}u_h^{i+1}&=L_h^{-1}l_h^{i+1}=L_h^{-1}[l_h^i-d(u_h^i)]\\&=L_h^{-1}[L_hu_h^i-(\tilde{L}_hu_h^i-\tilde{f}_h)].\end{aligned}\tag{7.5}$$

总之，无论是方法一对近似解作余量校正，还是方法二对离散方程右端项作余量校正，当 L_h 和 $\tilde{L}_h$ 都是线性离散算子时，(7.4)式和(7.5)式是一致的。所以就有**IDC 算法**

$$\begin{cases}u_h^0=L_h^{-1}f_h\\u_h^{i+1}=u_h^i-L_h^{-1}(\tilde{L}_hu_h^i-\tilde{f}_h),\quad i=1,2,\cdots.\end{cases}\tag{7.6}$$

当 L_h 和 $\tilde{L}_h$ 都是非线性离散算子时，人们可以利用(7.4)和(7.5)式获得相应的 IDC 算法[41~44]，(7.6)式仅对L_h线性时适用。本章只对 L_h 和 $\tilde{L}_h$ 都是线性的情形讨论之。

由(7.6)式可见，IDC 算法利用了(7.3)式，但避免求$\tilde{L}_h^{-1}$，这不仅避免了(7.3)式数值求解难及可能数值不稳定的缺点，而且发扬了 $\tilde{L}_h$ 具有高相容阶的长处，使得利用高精度残差向量 $\tilde{L}_hu_h^i-\tilde{f}_h$ 求校正量更有效果。IDC 算法也利用了(7.2)式，但只是充分发挥了 L_h 数值稳定的长处。IDC 算法可迂回达到数值求解(7.3)式的目的，因此 IDC 算法的思想是有启发性的。

从(7.6)式易知 IDC 算法的迭代矩阵是

$$T_h=I_h-L_h^{-1}\tilde{L}_h;\tag{7.7}$$

同样易知，若 L_h 和 $\tilde{L}_h$ 非奇异且

$$\rho(T_h)=\rho(I_h-L_h^{-1}\tilde{L}_h)=\rho(L_h^{-1}(L_h-\tilde{L}_h))<1,$$

则(7.6)式的解 u_h^i 收敛到(7.3)式的解 $\tilde{u}_h$。事实上，若迭代收敛，则当 $i\to\infty$时有 $u_h^{i+1}=u_h^i$，这样由(7.6)式推知 u_h^i 满足(7.3)式，即有 $u_h^i\to\tilde{u}_h$；另外，IDC 算法应要求 $\tilde{L}_h$ 非奇异，否则人们可选择 $v_h\neq0$ 使 $\tilde{L}_hv_h=0$，那么 $T_hv_h=v_h$，(7.6)式就不收敛了，这说明对于$\tilde{L}_h$非奇异性的要求是正当的。

§2 余量校正迭代误差估计

为了作误差估计，先简略介绍一些关于算子方程离散化的概念。

设(7.1)式的算子 L 是无穷维函数空间 $D\subset E\to\hat{D}\subset\hat{E}$ 的映射，为了数值计算的目的，人们按网格参数 h 将定解区域离散化，并将无穷维函数空间的算子和函数都限制到有限维空间中，从而得到离散方程

$$L^h u^h = f^h, \qquad (7.8)$$

其中，u^h 和 f^h 是网格节点处离散值构成的网格函数，它们分别属于有限维网格函数空间 E^h 和 $\hat{E}^h$。这样，人们很容易理解(7.1)和(7.8)式中诸映射之间的相互关系，见图7.3，其中 R_E^h 和 $\hat{R}_{\hat{E}}^h$ 是限制算子。

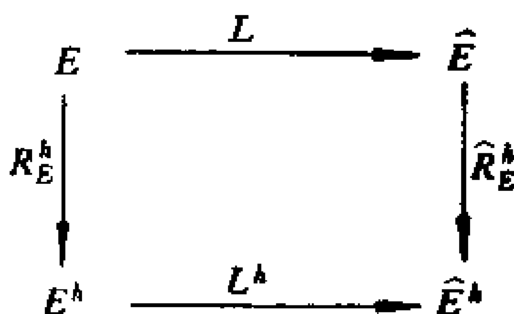


图 7.3

在无限维函数空间及其相应的有限维网格函数空间中定义的模量应满足下述关系：

$$\lim_{h\to 0}\|R_E^h u\|_{E^h} = \|u\|_E, \qquad \forall u\in E,$$

$$\lim_{h\to 0}\|\hat{R}_{\hat{E}}^h F\|_{\hat{E}^h} = \|f\|_{\hat{E}}, \qquad \forall E\in\hat{E}.$$

称离散方程(7.8)是数值稳定的，若存在与 u 无关的常数 c，使有

$$\|(L^h)^{-1}\|_{E^h}\leqslant c \qquad (7.9)$$

成立。这也是离散算子 L^h 规则性的描述。

设 u^* 和 $(u^h)^*$ 分别是(7.1)和(7.8)式的准确解，则 $L^h R_E^h u^* - \hat{R}_{\hat{E}}^h L u^*$ 反应了椭圆离散方程(7.8)对原方程(7.1)的局部离散误差或截断误差，因此椭圆离散方程(7.8)的 p 阶相容性条件可描述为

$$\|L^h R_E^h u^* - \hat{R}_E^h L u^*\|_{\hat{E}^h} = O(h^p) \quad (h\to 0),$$

或

$$\|L^h R_E^h - \hat{R}_E^h L\|_{\hat{E}^h} = O(h^p). \quad (h\to 0). \tag{7.10}$$

若离散算子 L^h 是规则的，即 L^h 满足(7.9)式，则由相容性条件(7.10)可推出差分解的收敛性，事实上当 L^h 是线性情形时，有

$$L^h R_E^h u^* - \hat{R}_E^h L u^* = L^h R_E^h u^* - \hat{R}_E^h f$$
$$= L^h R_E^h u^* - f^h = L^h(R_E^h u^* - (u^h)^*) = O(h^p),$$

所以

$$\|R_E^h u^* - (u^h)^*\|_{E^h} = O(h^p).$$

上述概念的数学描述同样适用于离散方程(7.2)和(7.3)，从而可进一步讨论 IDC 误差估计问题。为此，记限制算子

$$\bar{R}_h: E\to E^h, \quad R_h: \hat{E}\to\hat{E}^h, \quad \tilde{R}_h: \hat{E}\to\hat{E}^h.$$

那么(7.2)和(7.3)式的右端项可表示为

$$f_h = R_h f, \quad \tilde{f}_h = \tilde{R}_h f, \tag{7.11}$$

还记 $u_h^* = \bar{R}_h u$，其中 u 是(7.1)式的解。下面定理表明：余量校正迭代无需求解离散方程(7.3)，只要经有限步迭代，即可达到(7.3)式解所具有的数值解精度。

定理7.1 设 $\tilde{L}_h$ 非奇异且(7.11)式成立，还设下面三个条件成立：

1. L_h 是数值稳定的(参见(7.9)式)，即

$$\|L_h^{-1}\|_{E^h}\leqslant c; \tag{7.12}$$

2. L_h 和 $\tilde{L}_h$ 分别具有 κ 和 $\tilde{\kappa}$ $(\kappa<\tilde{\kappa})$ 相容阶(参见(7.10)式)，即

$$\|R_h L - L_h\bar{R}_h\|_{\hat{E}^h}\leqslant ch^{\kappa}, \tag{7.13}$$

$$\|\tilde{R}_h L - \tilde{L}_h\bar{R}_h\|_{\hat{E}^h}\leqslant ch^{\tilde{\kappa}}; \tag{7.14}$$

3. L_h 和 $\tilde{L}_h$ 是相对相容的，即

$$\|\tilde{L}_h - L_h\|_{\hat{E}^h}\leqslant ch^{\kappa}, \tag{7.15}$$

其中 $\|\tilde{L}_h\|_{\hat{E}^h} = ch^{\tilde{\kappa}}$，$\|L_h\|_{\hat{E}^h} = ch^{\kappa}$(参见第二章§2)；则余量校正

迭代法第 i 步迭代值 u_h^i 满足估计式：

$$\|u_h^i - u_h^*\|_{E^h} \leqslant c_i h^{\min(\tilde{\kappa}, (i+1)\kappa)} \|u\|_{E^*} \tag{7.16}$$

证明 由(7.6)式知 $u_h^0 - u_h^* = L_h^{-1}(R_h L - L_h R_h)u$，所以

$$\|u_h^0 - u_h^*\|_{E^h} = c h^{\kappa} \|u\|_E; \tag{7.17}$$

还知

$$u_h^i - u_h^* = L_h^{-1}(L_h - \tilde{L}_h)(u_h^{i-1} - u_h^*) + L_h^{-1}(\tilde{R}_h L - \tilde{L}_h \tilde{R}_h)u,$$

故利用(7.12)～(7.17)诸式和递推关系，就有

$$\begin{aligned}\|u_h^i - u_h^*\|_{E^h} &\leqslant c h^{\kappa} \|u_h^{i-1} - u_h^*\|_{E^h} + c h^{\tilde{\kappa}} \|u\|_E \\ &\leqslant c h^{\kappa}(c h^{\kappa} \|u_h^{i-2} - u_h^*\|_{E^h} + c h^{\tilde{\kappa}} \|u\|_E) + c h^{\tilde{\kappa}} \|u\|_E \\ &\cdots \\ &\leqslant c_i h^{\min(\tilde{\kappa}, (i+1)\kappa)} \|u\|_E. \quad \square\end{aligned}$$

(7.16) 式清楚地表明，IDC 方法的每个迭代步使近似解误差提高了 h^{κ} 阶，直至$(i+1)\kappa \geqslant \tilde{\kappa}$时，用 IDC 方法就能获得$\tilde{\kappa}$ 阶精度的数值解。因此，IDC 最佳迭代步数是 $i = \tilde{\kappa}/\kappa - 1$。

§3 余量校正多层网格迭代方法

本节的目的是将 IDC 方法和 MG 方法有效地结合起来，以形成更实用有效的算法。本节中所涉及的 MG 方法中的有关记号和意义都同于第四章。记网格层序列 $\{\Omega_l\}$，$l = 1, 2, \cdots, l$，h_l 是 Ω_l 的网格参数；在 Ω_l 层上定义的且与(7.2)和(7.3)式相应的离散方程分别记为

$$L_l u_l = f_l, \tag{7.18}$$

$$\tilde{L}_l \tilde{u}_l = \tilde{f}_l. \tag{7.19}$$

记 r 和 p 分别是两相邻网格层上网格函数空间之间的限制和内插算子；$S_l(L_l, u_l, d_l, j)$ 表示在 Ω_l 层上以 u_l 为初值，对离散方程 $L_l u_l = d_l$ 作 j 次光滑迭代；$\tilde{S}_l(\tilde{L}_l, u_l, \tilde{d}_l, j)$表示在$\Omega_l$

层上以u_l为初值，对离散方程$\tilde{L}_l\tilde{u}_l=\tilde{f}_l$作$i$次光滑迭代；用$MG(i,u_l,L_l,d_l,\nu)$表示在$\Omega_l,\cdots,\Omega_1$层上定义的以$u_l$为初始值且数值求解$L_lu_l=d_l$的$\nu$次多层网格($V$或$W$循环)迭代步。

正如§1中所述，在Ω_l层上利用(7.18)和(7.19)式的IDC方法共由2个部分组成：

$$\text{余量校正部分}\qquad u_l^{i+1}=u_l^i+v_l^i, \tag{7.20}$$

$$\text{迭代求解部分}\qquad L_lv_l^i=d(u_l^i)=\tilde{L}_lu_l^i-\tilde{f}_l. \tag{7.21}$$

因此，任一个有效的迭代方法都可用于求解(7.21)式，这样便可形成余量校正原则和其它迭代法组合而成的新的数值方法。本节将使用高效率的多层网格方法数值求解(7.21)式中的校正量，这就是本节要介绍的余量校正多层网格迭代方法，简记为DCMG。

算法DCMG1：

$$\begin{cases}\bar{u}_l:=u_l^i,\\ d_l:=\tilde{L}_l\bar{u}_l-\tilde{f}_l,\\ v_l:=MG(l,0,L_l,d_l,\nu),\\ u_l^{i+1}:=\bar{u}_l+v_l.\end{cases} \tag{7.22}$$

该算法的特点是只用MG方法的ν个循环迭代步求得近似校正量后，就作余量校正。因此，该算法不象IDC算法那样在有限步内完成计算工作，而是把余量校正过程当作一个无限的收敛过程。当然，DCMG1的解u_l^i和IDC的解$\bar{u}_l^i$收敛于同一个极限值$\tilde{u}_l$，即收敛于(7.3)式的准确解。

值得注意的是：若在DCMG1算法和IDC算法中取相同的初值$u_l^i=\bar{u}_l^i$，那么DCMG1算法经过若干（有限）步迭代后就能达到IDC算法的值$\bar{u}_l^{i+1}$，这是因为MG方法具有h无关的收敛率ξ的缘故，所以在DCMG1算法的每个迭代步中作了ν步MG循环迭代步后，有

$$\|u_l^{i+1}-\bar{u}_l^{i+1}\|\leqslant\xi^\nu\|u_l^i-\bar{u}_l^{i+1}\|=\xi^\nu\|\bar{u}_l^i-\bar{u}_l^{i+1}\|; \tag{7.23}$$

另外，IDC 算法的收敛率也可描述为

$$\|\bar{u}_l^{i+1}-\tilde{u}_l\|\leqslant\eta\|\bar{u}_l^i-\tilde{u}_l\|; \tag{7.24}$$

最后，利用(7.23)和(7.24)式可估计 DCMG1 算法的收敛率了。

定理7.2 设 ξ 和 η 分别是 MG 算法和 IDC 算法的收敛率，则由(7.22)式所描述的 DCMG1 算法的收敛率为

$$q\leqslant\eta+\xi^\gamma+\xi^\gamma\eta. \tag{7.25}$$

证明：利用(7.23)和(7.24)式有

$$\begin{aligned}\|u_l^{i+1}-\tilde{u}_l\|&\leqslant\|u_l^{i+1}-\bar{u}_l^{i+1}\|+\|\bar{u}_l^{i+1}-\tilde{u}_l\|\\&\leqslant\xi^\gamma\|\bar{u}_l^i-\bar{u}_l^{i+1}\|+\eta\|\bar{u}_l^i-\tilde{u}_l\|\\&\leqslant\xi^\gamma(\|\bar{u}_l^i-\tilde{u}_l\|+\|\bar{u}_l^{i+1}-\tilde{u}_l\|)+\eta\|\bar{u}_l^i-\tilde{u}_l\|\\&\leqslant(\eta+\xi^\gamma+\xi^\gamma\eta)\|\bar{u}_l^i-\tilde{u}_l\|.\ \square\end{aligned}$$

由此可见，对于给定的离散方程(7.2)和(7.3)，可利用(7.17)式估算出 η 的数值，若再知道 MG 每个循环迭代步的收敛率 ξ，则可估算出 DCMG1 算法的收敛率 q；只要适当增加 MG 循环迭代步数 ν，使 $q=\eta+\xi^\gamma+\xi^\gamma\eta<1$，就可保证DCMG1 算法收敛；特别地，当 MG 循环迭代步数足够大并足以达到准确求解 $L_lv_l=d_l$ 的情形时，就有 DCMG1 算法等效于IDC 算法的结果。

为了更好地发挥 MG 方法中粗网格校正过程的效果，可在 DCMG1 算法中增加预先光滑步骤，这样就有算法DCMG2：

$$\begin{cases}\bar{u}_l:=\tilde{S}_l(\tilde{L}_l,\ u_l^i,\ \tilde{f}_l,\ j),\\ d_l:=\tilde{L}_l\bar{u}_l-\tilde{f}_l,\\ v_l:=\mathrm{MG}(l,0,L_l,d_l,\nu),\\ u_l^{i+1}:=\bar{u}_l+v_l.\end{cases} \tag{7.26}$$

显然，用该算法获得的数值解 u_l^i 的极限也是(7.3)式的准确解 $\tilde{u}_l$。

下面推导 DCMG2 算法的迭代矩阵。设 $\tilde{S}_l$ 是该算法使用的

光滑迭代矩阵，则有

$$\bar{u}_l = u_l^j - (I - \tilde{S}_l^j)\tilde{L}_l^{-1}(\tilde{L}_l u_l^j - \tilde{f}_l),$$

再设 MG 方法的迭代矩阵是 M_l，则当初值为零时（具体参见第 4 章§4.1）有

$$v_l = M_l \cdot 0 - (I - M_l^\nu) L_l^{-1}(\tilde{L}_l \bar{u}_l - \tilde{f}_l);$$

利用这两个式子有

$$\begin{aligned} u_l^{j+1} &= \bar{u}_l + v_l = \bar{u}_l - (I - M_l^\nu) L_l^{-1}(\tilde{L}_l \bar{u}_l - \tilde{f}_l) \\ &= [I - (I - M_l^\nu) L_l^{-1} \tilde{L}_l] \bar{u}_l + (I - M_l^\nu) L_l^{-1} \tilde{f}_l, \end{aligned}$$

总之可直接写出 DCMG2 算法的迭代矩阵

$$\begin{aligned} T_{l2} &= [I - (I - M_l^\nu) L_l^{-1} \tilde{L}_l] \tilde{S}_l^j \\ &= (I - L_l^{-1}\tilde{L}_l)\tilde{S}_l^j + M_l^\nu L_l^{-1}\tilde{L}_l \tilde{S}_l^j \\ &= (\tilde{L}_l^{-1} - L_l^{-1})\tilde{L}_l \tilde{S}_l^j + M_l^\nu L_l^{-1} \tilde{L}_l \tilde{S}_l^j. \end{aligned} \tag{7.27}$$

关于 DCMG2 算法的收敛性估计问题，只要利用第 4 章 §4 中的办法，可估计出下面诸式：

$$\|\tilde{L}_l^{-1} - L_l^{-1}\| \leqslant c_1 h_l^{-\tilde{r}}, \quad \|\tilde{L}_l \tilde{S}_l^j\| \leqslant \varepsilon_1(j) h^{\tilde{r}},$$

$$\|M_l^\nu\| \leqslant \xi^\nu < 1, \quad \|L_l^{-1}\tilde{L}_l \tilde{S}_l^j\| \leqslant \varepsilon_2(j),$$

其中 $\varepsilon_1(j) \to 0$，$\varepsilon_2(j) \to 0$，$j \to \infty$，就可获得

$$\|T_{l2}\| \leqslant c_1 \varepsilon_1(j) + \varepsilon_2(j)\xi^\nu.$$

因此，只要适当增加光滑次数 j，使 $\varepsilon_1(j)$和$\varepsilon_2(j)$足够小且适当增加MG循环迭代步数 ν，使 ξ^ν 足够小，就能保证算法DCMG2收敛。

读者不难发现，在算法 DCMG2 中预先光滑步和 MG 循环中的光滑步是针对不同的离散方程实现的，因此，一个自然的改变办法是将这两种光滑算子统一起来。算法 DCMG3:

$$\begin{cases} \bar{u}_l := S_l(L_l, u_l^j, f_l, j), \\ d_l := \tilde{L}_l \bar{u}_l - \tilde{f}_l, \\ v_l := \mathrm{MG}(l, 0, L_l, d_l, \nu), \\ u_l^{j+1} := \bar{u}_l + v_l, \end{cases} \tag{7.28}$$

显然，在该算法中，预先光滑过程和求校正量的MG过程是分开书写的，当然也可以把这两部分过程统一于同一个MG迭代程序中书写。无论如何书写，用 DCMG3 算法与用通常的 MG 算法求解过程的不同之处仅在于 d_l，前者 $d_l := \tilde{L}_l\bar{u}_l - \tilde{f}_l$，后者 $d_l := L_l\bar{u}_l - f_l$。

仿(7.27)式，可直接写出算法DCMG3 的迭代计算式：

$$
\begin{aligned}
u_l^{j+1} &= \bar{u}_l - (I - M_l^\nu)L_l^{-1}(\tilde{L}_l\bar{u}_l - \tilde{f}_l) \\
&= u_l^j - (I - S_l^j)L_l^{-1}(L_l u_l^j - f_l) \\
&\quad - (I - M_l^\nu)L_l^{-1}[\tilde{L}_l(u_l^j - (I - S_l^j)L_l^{-1}(L_l u_l^j - f_l)) \\
&\qquad - \tilde{f}_l],
\end{aligned}
\tag{7.29}
$$

和迭代矩阵

$$
T_{l3} = [I - (I - M_l^\nu)L_l^{-1}\tilde{L}_l]S_l^j. \tag{7.30}
$$

读者不难明白，算法 DCMG3 解的极限不是 (7.2)或 (7.3)式的准确解，而是另一个值 $\bar{u}_l^*$，但 $\bar{u}_l^*$ 和 $\hat{u}_l$((7.3)式的解) 相差很小。

事实上，DCMG3 算法所得的极限值 u_l^* 满足(7.29)式，即有

$$
\begin{aligned}
\bar{u}_l^* &= \bar{u}_l^* - (I - S_l^j)L_l^{-1}(L_l\bar{u}_l^* - f_l) \\
&\quad - (I - M_l^\nu)L_l^{-1}[\tilde{L}_l(\bar{u}_l^* - (I - S_l^j)L_l^{-1}(L_l\bar{u}_l^* - f_l)) - \tilde{f}_l];
\end{aligned}
$$

$\hat{u}_l$ 满足恒等式：

$$
\begin{aligned}
\hat{u}_l &= \hat{u}_l - (I - S_l^j)L_l^{-1}(L_l\hat{u}_l - f_l) \\
&\quad - (I - M_l^\nu)L_l^{-1}[\tilde{L}_l(\hat{u}_l - (I - S_l^j)L_l^{-1}(L_l\hat{u}_l - f_l)) - \tilde{f}_l] \\
&\quad + [I - (I - M_l^\nu)L_l^{-1}\tilde{L}_l](I - S_l^j)L_l^{-1}(L_l\hat{u}_l - f_l);
\end{aligned}
$$

若记 $E_l^* = \bar{u}_l^* - \hat{u}_l$，则由上两式相减得

$$
E_l^* = T_{l3}E_l^* - [I - (I - M_l^\nu)L_l^{-1}\tilde{L}_l](I - S_l^j)L_l^{-1}(L_l\hat{u}_l - f_l),
$$

即

$$
\begin{aligned}
E_l^* = -(I - T_{l3})^{-1}[I - (I - M_l^\nu)L_l^{-1}\tilde{L}_l](I \\
- S_l^j)L_l^{-1}(L_l\hat{u}_l - f_l).
\end{aligned}
\tag{7.31}
$$

现对(7.31)式进行估计。因为 $L_l\tilde{u}_l - f_l$ 近似描述了(7.2)式的相容逼近误差，所以，

$$\|L_l\tilde{u}_l - f_l\| = O(h^{\kappa});$$

又因为对确定的光滑方法，例如 GS 方法，光滑算子 S_l 的模量一般是与派生出此光滑算子的离散算子 L_l 具有相同的阶，具体地可参见第二章§1，所以有

$$\|(I - S_l^j)L_l^{-1}\| = O(h^{\kappa}), \qquad j \geqslant 1;$$

还注意到(7.31)式前 2 个因子都是常数阶；故由(7.31)式可知，当算法 DCMG3 收敛时($\|T_{l3}\| < 1$) 它的解与 (7.3) 式的解偏离很小，比 $O(h^{2\kappa})$量级还要小。

§4 算例与其它

本节考虑一个简单的数值例子：

$$\begin{cases}-\Delta u + (1 + x^2 + y^2)u = g(x, y), & (x, y) \in (0,1)^2 = G,\\ u = 0, & (x, y) \in \partial G,\end{cases}$$

其真解是

$$u(x, y) = \sin(\pi x)\sin(\pi y) + 0.2\sin(5\pi x)\sin(5\pi y).$$

取 $h = 1/32$ 的网格分划，并作出 5 点格式(二阶精度)和 9 点格式(四阶精度)。

先用确定的完全多层网格方法程序（层间光滑一律用 2 次 GS迭代，在各个中间层上都作一次V循环)求解 5 点格式离散方程，获得 $O(h^2)$阶精度的数值解，数值解与真解之误差最大模是 $E = 0.4797 \times 10^{-2}$，所用时间记为 T；再用这个确定的完全多层网格方法程序与余量校正法结合，形成DCMG1、DCMG2和DCMG3并计算本节的例子；我们将计算结果列于表7.1中。

表 7.1 的"精度"为数值解与真解误差的最大模；"效率"是个常数，它是用余量校正多层网格方法计算时间与这个确定的完全多层网格方法求解 5 点格式所用时间 T 的比值；表 7.1 中

NCYCL表示两次相邻余量校正步之间所完成的多层网格循环迭代步；表 7.1 中的算法 DCMG1 没有用到细网层上的光滑过程，算法 DCMG2 的细网层光滑采用收敛率为 5/8 的 Jacobi 松弛 2 个迭代步，算法 DCMG3 的细网层光滑采用收敛率为 1/2 的 Jacobi 松弛 2 个迭代步。

表7.1 余量校正多层网格迭代法的求解精度与效率($h=1/32$)

NCYCL	DCMG1		DCMG2		DCMG3	
	精度	效率	精度	效率	精度	效率
1	.0980$\bar{E}$	1.33	.0078$\bar{E}$	1.52	.0097$\bar{E}$	1.47
2	.0160$\bar{E}$	1.71	.0066$\bar{E}$	2.10	.0077$\bar{E}$	2.00
3	.0070$\bar{E}$	2.10	.0067$\bar{E}$	2.78	.0076$\bar{E}$	2.43

表 7.1 说明了如下几点事实：

1．余量校正多层网格迭代方法以较少的计算量获得较高的数值解精度。表中数值已表明，DCMG1、DCMG2和DCMG3算法的数值解精度都已达$O(h^4)$量级，花费的计算时间约为$1.33T\sim2.78T$；但若将网格进一步细化为$h=1/64$，则求解相应的 5 点离格式将花费时间 $4T$ 且只能达到$0.25\bar{E}$的精度。

2．虽然算法DCMG1、DCMG2和DCMG3的极限点不同，但数值解精度都能达到$O(h^4)$量级。DCMG3 算法不但简单易行，而且效率也高，所以它是较好的。

3．细网层作适当光滑是会产生良好结果的。没有细层光滑的算法 DCMG1 的效果就比有细层光滑的算法 DCMG2 和 DCMG3 的效果差些。当然，过多的细层光滑步及两相邻余量校正步之间采用过多的 MG 循环迭代步都会浪费计算工作量。

作为本章结束，我们进一步讨论扩展应用余量校正迭代方法

的问题。正如§3中所述，余量校正迭代法由两个基本部分组成，即余量校正部分(7.20)式和迭代求解部分(7.21)式，因此研究改进余量校正迭代方法有3个主要方面。

第一方面是想法设计构造高精度的离散格式。

第二方面是提高余量校正的效率。例如，L_h^1、L_h^2和$\tilde{L}_h$都是L的合理逼近，它们的相容逼近阶分别是κ_1、κ_2和$\tilde{\kappa}$且$\kappa_1<\tilde{\kappa}$，$\kappa_2<\tilde{\kappa}$，分别用L_h^1和L_h^2作余量校正迭代，据(7.6)式有

$$\begin{cases} u_h^{i+1/2}=(I-(L_h^1)^{-1}\tilde{L}_h)u_h^i+(L_h^1)^{-1}\tilde{f}_h, \\ u_h^{i+1}=(I-(L_h^2)^{-1}\tilde{L}_h)u_h^{i+1/2}+(L_h^2)^{-1}\tilde{f}_h; \end{cases}$$

这样组合而成的余量校正迭代步

$$u_h^{i+1}=[I-(L_h^2)^{-1}\tilde{L}_h][I-(L_h^1)^{-1}\tilde{L}_h]u_h^i \\ +[(I-(L_h^2)^{-1}\tilde{L}_h)(L_h^1)^{-1}+(L_h^2)^{-1}]\tilde{f}_h$$

就有整体迭代矩阵

$$T_h=[I-(L_h^2)^{-1}\tilde{L}_h][I-(L_h^1)^{-1}\tilde{L}_h];$$

这种多步余量校正迭代过程可能会产生好的效果。

第三方面是加速迭代求解过程，将各种高效率迭代法与余量校正原则组合成新的有效算法。例如，§3中余量校正多层网格迭代方法就是一例，这种结合不仅可以简单引用现成的MG子程序，而且进一步利用了MG方法求解(7.21)式的高效率优点。又例如，将第五章介绍的逐层子空间迭代法、第六章介绍的逐层分裂迭代法、第八章介绍的缩减方法等与余量校正迭代原则相结合，这定将产生高效率高精度的数值求解结果。这里介绍三种**逐层余量校正迭代法**(Iterative Defect Correct Methods in Successive Level)，简记为IDCSL**方法**。算法IDCSL1：

$$\begin{cases} u_l^0=L_l^{-1}f_l, \\ d_l=\tilde{L}_lu_l^i-\tilde{f}_l, \\ d_k=I_l^kd_l,\ k=1,2,\cdots,l, \\ v_1=\mathrm{relax}(L_1,0,d_1,\nu_1), \end{cases}$$

$$\begin{cases} \bar{v}_k = I_{k-1}^k v_{k-1}, \\ v_k = \mathrm{relax}(L_k, \bar{v}_k, d_k, \nu_k), \end{cases} \Big\} k = 2, \cdots, l, \\ u_l^{i+1} = u_l^i + v_l, \quad i = 0, 1, \cdots,$$

其中 relax(L_k, $\bar{v}_k$, d_k, ν_k) 表示在 Ω_k 网格层上对离散方程 $L_k w_k = d_k$ 用初值 $\bar{v}_k$ 作 ν_k 次松弛迭代，当k=1时，则表示在 Ω_1 上准确求解离散方程 $L_1 w_1 = d_1$，下同。

算法IDCSL2：

$$\tilde{u}_k = \mathrm{IDC}(L_k,\ \tilde{L}_k,\ \bar{u}_k,\ f_k,\ \tilde{f}_k), \quad k = 1, 2, \cdots, l,$$

其中，$\bar{u}_1 = L_1^{-1} f_1$，$\bar{u}_k = I_{k-1}^k \tilde{u}_{k-1}$，IDC($L_k$, $\tilde{L}_k$, $\bar{u}_k$, f_k, $\tilde{f}_k$) 表示利用离散方程 $L_k w_k = f_k$、$\tilde{L}_k \tilde{w}_k = \tilde{f}_k$ 和初值 $\bar{u}_k$ 作余量校正迭代。

算法IDCSL3：

$$\begin{cases} u_l^0 = L_l^{-1} f_l, \quad u_k^i = I_l^k u_l^i, \quad k = 1, 2, \cdots, l-1, \\ d_1 = \tilde{L}_1 u_1^i - \tilde{f}_1, \quad v_1 = L_1^{-1} d_1, \\ \left.\begin{array}{l} \bar{v}_k = I_{k-1}^k v_{k-1}, \quad \bar{u}_k = u_k^i - \bar{v}_k, \\ d_k = \tilde{L}_k \bar{u}_k - \tilde{f}_k, \\ v_k = \mathrm{relax}(L_k, d_k, 0, \nu_k), \end{array}\right\} k = 1, \cdots, l, \\ u_l^{i+1} = v_l^i - v_l, \quad i = 0, 1, \cdots. \end{cases}$$

最后我们指出，余量校正迭代方法是实用的，而且该方法为解决用线性问题逼近非线性问题、用简单问题逼近复杂问题、用低精度问题逼近高精度问题和用稳定问题逼近稳定性不够好的问题提供了一种有效途径。

第八章　缩减方法

缩减方法是 J.Schröder 和 U.Trottenberg 在 70 年代初开始研究的，该方法的基本思想仍然是将细网格层上的数值求解问题转化为粗网格层上的数值求解问题；该方法的实施办法是在离散微分方程两端同时作用一个“**缩减算子**”；这样，粗网格离散微分方程的未知数数目被缩减为原来的一半，而且粗网格上的数值解保持了细网格上离散微分方程数值解精度的要求；该方法是快速且数值稳定的。

§1　解常微分方程边值问题的缩减方法

首先考虑简单的模型问题

$$\begin{cases} -u''(x)=f(x), & x\in(0,\ 1), \\ u(0)=a, & u(1)=b. \end{cases} \tag{8.1}$$

将$[0,\ 1]$ n 等分，$n=2^p$，$p\in N$，$h=\frac{1}{n}$。用差分方法或线性有限元方法即可得到大家熟知的离散方程

$$[-1\quad 2\quad -1]_h u(x)=r(x),\quad x\in\Omega_0, \tag{8.1E_0}$$

其中，$u(0)=a$，$u(1)=b$，$r(x)=h^2 f(x)$，

$$\Omega_0:=\{x\,|\,x=ih,\ i=1,\ 2,\ \cdots,\ n-1\},$$

$$[\alpha\ \beta\ \gamma]_h u(x):=\alpha u(x-h)+\beta u(x)+\gamma u(x+h).$$

为了快速求解细网格层 Ω_0 上的离散方程(E_0)，可将(E_0)变换到较粗的网格层 Ω_1 $(h_1=2h)$上。为此，于(E_0)两端左乘一个“缩减算子”$[1\ 2\ 1]_h$，就有

$$[1\ 2\ 1]_h[-1\ 2\ -1]_h u(x)=[1\ 2\ 1]_h r(x),\quad x\in\Omega_0;$$

按局部离散算子的运算规则

$$[a\ b\ c]_h[\alpha\ \beta\ \gamma]_h u(x):=[a\ b\ c]_h \alpha u(x-h)$$

$$+[a\ \ b\ \ c]_h\beta u(x)+[a\ \ b\ \ c]_h\gamma u(x+h),$$

有

$$[-1\ \ 0\ \ 2\ \ 0\ \ -1]_h u(x)=[1\ \ 2\ \ 1]_h r(x),\quad x\in\Omega_0;$$

从而得到第一次**缩减方程**

$$[-1\ \ 2\ \ -1]_{2h}u(x)=r_1(x),\qquad x\in\Omega_1,\tag{8.1E_1}$$

其中，$r_1(x)=[1\ \ 2\ \ 1]_h r(x)$，

$$\Omega_1:=\left\{x\,|\,x=i2h,\ i=1,\ 2,\cdots,\ \frac{n}{2}-1\right\}.$$

再在(E_1)式两边同乘缩减算子$[1\ \ 2\ \ 1]_{2h}$就得到第二次缩减方程

$$[-1\ \ 2\ \ -1]_{4h}u(x)=r_2(x),\qquad x\in\Omega_2.\tag{8.1E_2}$$

一般地，μ次缩减方程为

$$[-1\ \ 2\ \ -1]_{h(\mu)}u(x)=r_\mu(x),\qquad x\in\Omega_\mu,\tag{8.1E_μ}$$

其中，

$$r_\mu(x)=[1\ \ 2\ \ 1]_{h(\mu-1)}r_{\mu-1}(x),\qquad h(\mu)=h_\mu,$$

$$\Omega_\mu=\{x\,|\,x=ih_\mu,\ i=1,\ 2,\cdots,n_\mu-1\},$$

$$n_\mu=h_\mu^{-1},\ h_\mu=2^\mu h,\ \mu=1,\ 2,\cdots,p-1.$$

注意，方程(E_{p-1})只是关于一个未知量$u\left(\frac{1}{2}\right)$的方程，可直接算出$u\left(\frac{1}{2}\right)$；方程$(E_{p-2})$是关于$u\left(\frac{1}{4}\right)$、$u\left(\frac{1}{2}\right)$和$u\left(\frac{3}{4}\right)$的方程组，由于$u\left(\frac{1}{2}\right)$已算出，所以$u\left(\frac{1}{4}\right)$和$u\left(\frac{3}{4}\right)$也可以从方程$(E_{p-2})$中直接计算得到，并不用解什么方程组；如此下去，就可以计算出Ω_0上所有节点处的数值解。

具体地，将[0，1]分成2^3等份，有7个内点，(E_0)式可经2次缩减依次得(E_1)式和(E_2)式，其形式为

$$[-1\ \ 2\ \ -1]_{2h}u(x)=r_1(x),\tag{E_1}$$

$$[-1\ \ 2\ \ -1]_{4h}u(x)=r_2(x),\tag{E_2}$$

$$\begin{aligned}r_1(x)&=[1\ \ 2\ \ 1]_h r(x)\\&=h^2[f(x-h)+2f(x)+f(x+h)];\end{aligned}$$

$$
\begin{aligned}
r_2(x) &= [1\ \ 2\ \ 1]_{2h} r_1(x) \\
&= r_1(x-2h) + 2r_1(x) + r_1(x+2h) \\
&= h^2[f(x-3h) + 2f(x-2h) + 3f(x-h) + 4f(x) \\
&\quad + 3f(x+h) + 2f(x+2h) + f(x+3h)].
\end{aligned}
$$

所以(E_2)式的最后形式为

$$
-u_0 + 2u_4 - u_8 = r_2(x_4),
$$

从而很容易求得 u_4。同样从(E_1)式中很容易求得 u_2 和 u_6，再从(E_0)式中直接计算出 u_1、u_3 和 u_7。

缩减方法可方便地应用于变系数两点边值问题

$$
\begin{cases}
-(p(x)u'(x))' + q(x)u(x) = f(x), & x \in (0,1), \\
u(0) = a, \ u(1) = b.
\end{cases}
\tag{8.2}
$$

当 $n = 2^p$，$x_i = ih$ 时，离散方程为

$$
\begin{cases}
\left[-p\left(x_i - \frac{h}{2}\right) \quad p\left(x_i - \frac{h}{2}\right) + p\left(x_i + \frac{h}{2}\right) \right. \\
\left. + h^2 q(x_i) \quad -p\left(x_i + \frac{h}{2}\right) \right]_h u(x_i) = h^2 f(x_i), \\
u(0) = a, \ u(1) = b, \quad i = 1, 2, \cdots, n-1.
\end{cases}
\tag{8.3}
$$

数值求解 (8.3) 时，如 h 足够小，那么由于舍入误差影响，将有 $p\left(x_i - \frac{h}{2}\right) + p\left(x_i + \frac{h}{2}\right) + h^2 q(x_i)$和 $p\left(x_i - \frac{h}{2}\right) + p\left(x_i + \frac{h}{2}\right)$ 的数值相等，这样(8.2)式中的$q(x)u(x)$的影响就会失掉，使数值解出现某些不真实情形.为适当顾及这一问题且适当考虑(8.2)式系数舍入误差对数值求解 (8.3) 式稳定性的影响，可按下述步骤实现缩减过程。为此将(8.3)式记为

$$
\begin{gathered}
[-a_i \quad a_i + b_i + c_i \quad -c_i]_h u(x_i) = r_i, \\
i = 1, 2, \cdots, n-1.
\end{gathered}
\tag{8.3E_0}
$$

将每三个相邻方程作合适的线性组合，可得到关于 $x_i \in \Omega_1$ ($h_1 = 2h$) 的第一次缩减方程

$$
[-A_i \quad A_i + B_i + C_i \quad -C_i]_{2h} u(x_i) = R_i,
\tag{8.3E_1}
$$

表8.1 数值求解二点边值问题的追赶方法是数值稳定的

n \ d_n \ k	$k=1$	$k=2$	$k=5$	$k=10$	$k=20$	$k=50$	$k=100$
4	0.37(−2)	0.39(−2)	0.56(−1)	0.33(−2)	0.35(−5)	0.11(−3)	0.82(−4)
8	0.87(−3)	0.19(−2)	0.11(−1)	0.28(−1)	0.16(−2)	0.51(−2)	0.55(−4)
16	0.22(−3)	0.43(−3)	0.13(−2)	0.56(−2)	0.14(−1)	0.85(−2)	0.25(−2)
32	0.56(−4)	0.11(−3)	0.29(−3)	0.64(−3)	0.28(−2)	0.10(−1)	0.43(−2)
64	0.14(−4)	0.28(−4)	0.71(−4)	0.14(−3)	0.32(−3)	0.29(−2)	0.52(−2)
128	0.39(−5)	0.69(−5)	0.18(−4)	0.36(−4)	0.72(−4)	0.22(−3)	0.15(−2)
256	0.15(−5)	0.20(−5)	0.52(−5)	0.95(−5)	0.19(−4)	0.46(−4)	0.11(−3)
512	0.19(−5)	0.19(−5)	0.33(−5)	0.38(−5)	0.52(−5)	0.12(−4)	0.24(−4)
1024	0.19(−5)	0.29(−5)	0.29(−5)	0.29(−5)	0.29(−5)	0.43(−5)	0.76(−5)
2048	0.24(−5)	0.33(−5)	0.29(−5)	0.33(−5)	0.29(−5)	0.24(−5)	0.29(−5)
4096	0.29(−5)	0.43(−5)	0.38(−5)	0.33(−5)	0.33(−5)	0.24(−5)	0.29(−5)
8192	0.29(−5)	0.33(−5)	0.29(−5)	0.33(−5)	0.33(−5)	0.43(−5)	0.29(−5)
16384	0.43(−5)	0.48(−5)	0.48(−5)	0.38(−5)	0.43(−5)	0.38(−5)	0.48(−5)
32768	0.43(−5)	0.43(−5)	0.48(−5)	0.67(−5)	0.38(−5)	0.38(−5)	0.48(−5)

$$i=2j,\ j=1,\ 2,\cdots,\frac{n}{2}-1,$$

其中$A_i=\alpha_i a_{i-1}$, $\alpha_i=a_i(a_{i-1}+b_{i-1}+c_{i-1})^{-1}$,

$C_i=\nu_i c_{i+1}$, $\nu_i=c_i(a_{i+1}+b_{i+1}+c_{i+1})^{-1}$,

$B_i=\alpha_i b_{i-1}+b_i+\nu_i b_{i+1}$, $R_i=\alpha_i r_{i-1}+r_i+\nu_i r_{i+1}$.

用相同的办法可得 (E_μ) $(\mu=1,\ 2,\cdots,p-1)$。不难看出，系数B_i是分三项求和计算的，这对数值稳定是有利的。

上述缩减方法的计算量$O(N)$是最佳的，N 是细网格层中网格节点个数。缩减方法是数值稳定的。

例 用缩减法计算边值问题：

$$\begin{cases}-((2+\cos\pi x)u'(x))'+2k^2\pi^2(1-\cos k\pi x)u(x)=f(x),\\ f(x)=2k\pi(2k\pi x(1-\cos k\pi x)-\sin k\pi x),\\ u(0)=0,\ u(1)=2,\end{cases}$$

其真解为$u=2x-(k\pi)^{-1}\sin k\pi x$。数值解的相对误差$\delta_n=|u-u_n|/|u|$列于表8.1中。从表 8.1 中的数值可见，由于真解和边值问题的系数是振荡的，所以数值解随着区域等分数n的增长而出现 3 种情况：当n太小时，数值解无规则；当n增加时，数值解体现了$O(h^2)$阶的精度，这一部分数值已在表中用长线条划出；n继续增加，精度无改善，这是由实际计算的舍入误差所致，但相对误差的精度并不变坏，这表明了缩减方法对舍入误差不敏感，缩减方法是数值稳定的。

§2 解偏微分方程的缩减方法

缩减方法也可有效地用于数值求解偏微分方程。为此先研究局部离散格式及其运算的规律，再研究缩减方法的构造和实施。

2.1 局部离散格式及其运算规律

考虑矩形网格

$$G_h=\{P\in R^2\mid P=P_0+P_{ij},\ i、j=0,\ \pm1,\ \pm2,\cdots\} \quad (8.4)$$

的情形，其中P_0是网格中的某个固定点，P_{ij}的坐标是$x=ih$，$y=jh$，h是网格步长，P按坐标式编号。

局部离散格式正如第一章定义的，表为

$$[S_k]_h := \begin{pmatrix} & & \vdots & \vdots & \vdots & & \\ & & s_{-1,1} & s_{0,1} & s_{1,1} & & \\ \cdots & s_{-2,0} & s_{-1,0} & s_{0,0} & s_{1,0} & s_{2,0} & \cdots \\ & & s_{-1,-1} & s_{0,-1} & s_{1,-1} & & \\ & & \vdots & \vdots & \vdots & & \end{pmatrix},$$

设局部离散格式$[S_k]_h$的系数s_{ij}全是实数，且中心对称，$s_{0,0}$是对称中心，s_{ij}只是有限个非零。

局部离散格式$[S_k]_h$作用于P点处的定义可表示为：

$$[S_k]_h u(P) := \sum_{i,j=-\infty}^{+\infty} s_{ij} u(P+P_{ij}). \tag{8.5}$$

若设$[T_k]_h$也是一个局部离散格式，则$([S_k]_h+[T_k]_h)$和$a[S_k]_h$是通常意义下的局部离散格式(参见§1)；$[S_k]_h\cdot[T_k]_h=[T_k]_h\cdot[S_k]_h$也是局部离散格式(参见§1)，其系数可表示为

$$g_{ij}=\sum_{\mu,\nu=-\infty}^{+\infty} s_{\mu,\nu}\, t_{i-\mu,j-\nu}. \tag{8.6}$$

为便于局部离散格式的构造和运算，定义偶的、奇的和共轭的局部离散格式。对于局部离散格式$[S_k]_h$来说，如果$i+j$等于奇数时有其元素$s_{ij}=0$，则称这个局部离散格式是偶的；如果当$i+j$等于偶数时有其元素$s_{ij}=0$，则称这个局部离散格式是奇的。利用局部离散格式相乘的定义易知，若$[S_k]_h$和$[T_k]_h$都是奇的或都是偶的，则$[S_k]_h\cdot[T_k]_h$或$[T_k]_h\cdot[S_k]_h$是偶的，否则就是奇的。若记$[S_k]_h^+$、$[S_k]_h^-$和$[\bar{S}_k]_h$分别为相应于$[S_k]_h$的**偶的、奇的和共轭的局部离散格式**，则有下列关系成立：

$$\begin{cases} [S_k]_h=[S_k]_h^+ - [S_k]_h^-, \\ [\bar{S}_k]_h=[S_k]_h^+ + [S_k]_h^- \quad ([\bar{S}_k]_h \text{ 的系数是 } (-1)^{i+j}s_{ij}), \\ [\bar{S}_k]_h[S_k]_h=([S_k]_h^+)^2-([S_k]_h^-)^2 \text{ 是偶的离散格式。} \end{cases} \tag{8.7}$$

例如5点差分格式$[S_k]_h$，按(8.5)式和(8.7)式有

$$\begin{cases} [S_k]_h=\begin{pmatrix} & -1 & \\ -1 & 4 & -1 \\ & -1 & \end{pmatrix},\quad [\bar{S}_k]_h=\begin{pmatrix} & 1 & \\ 1 & 4 & 1 \\ & 1 & \end{pmatrix},\quad [S_k]_h^+=\begin{pmatrix} & 0 & \\ 0 & 4 & 0 \\ & 0 & \end{pmatrix}, \\ [S_k]_h^-=\begin{pmatrix} & 1 & \\ 1 & 0 & 1 \\ & 1 & \end{pmatrix},\quad [\bar{S}_k]_h[S_k]_h=\begin{pmatrix} & & -1 & & \\ & -2 & 0 & -2 & \\ -1 & 0 & 12 & 0 & -1 \\ & -2 & 0 & -2 & \\ & & -1 & & \end{pmatrix}. \end{cases} \tag{8.8}$$

又如9点差分格式，有

$$\begin{cases} [S_k]_h=\begin{pmatrix} -1 & -4 & -1 \\ -4 & 20 & -4 \\ -1 & -4 & -1 \end{pmatrix},\quad [\bar{S}_k]_h=\begin{pmatrix} -1 & 4 & -1 \\ 4 & 20 & 4 \\ -1 & 4 & -1 \end{pmatrix}, \\ [S_k]_h^+=\begin{pmatrix} -1 & 0 & -1 \\ 0 & 20 & 0 \\ -1 & 0 & -1 \end{pmatrix},\quad [S_k]_h^-=\begin{pmatrix} 0 & 4 & 0 \\ 4 & 0 & 4 \\ 0 & 4 & 0 \end{pmatrix}, \\ [\bar{S}_k]_h[S_k]_h=\begin{pmatrix} 1 & 0 & -14 & 0 & 1 \\ 0 & -72 & 0 & -72 & 0 \\ -14 & 0 & 340 & 0 & -14 \\ 0 & -72 & 0 & -72 & 0 \\ 1 & 0 & -14 & 0 & 1 \end{pmatrix}. \end{cases} \tag{8.9}$$

下面进一步观察局部离散格式在红黑(RB)分划网格层序列$\{G_k\}$中的表现。G_0就是(8.4)式中的G_h，G_k可表示为

$$G_k=\{P\in R^2 \mid P=P_0+P_{ij}^k;\ i,\ j=0,\ \pm1,\ \pm2,\cdots\}, \tag{8.10}$$

对于确定的P_0，可通过P_{ij}^k坐标的选择来确定不同网格步长的G_k。具体地，P_{ij}^k坐标的选择办法是

$$\text{若}\begin{Bmatrix} k\ \text{为偶数} \\ k\ \text{为奇数} \end{Bmatrix},\ \text{则}P_{ij}^k\text{的坐标为}\begin{Bmatrix} x=ih_k,\ y=jh_k \\ \xi=ih_k,\ \eta=jh_k \end{Bmatrix},$$

其中$h_k=2^{k/2}h$，$\xi=(1/2)\sqrt{2}(x+y)$，$\eta=(1/2)\sqrt{2}(y-x)$。换

句话说，G_0是如(8.4)式所示的矩形网格，其网格步长为h_0；G_1的网格步长为$h_1=\sqrt{2}\,h_0$，G_1是G_0旋转$\pi/4$并将步长变为h_1的结果；再将G_1旋转$\pi/4$且令$h_2=\sqrt{2}\,h_1$；….图8.1表明了$P_0=(0,0)$，$h_0=a/8$的网格G_0和$G_k(k=1,2,3,4)$的关系，图中大的编号节点属于小编号的网格层，即$G_{k+1}\subset G_k$ (k = 0, 1, 2, …)．详细情形还可参见图8.3．

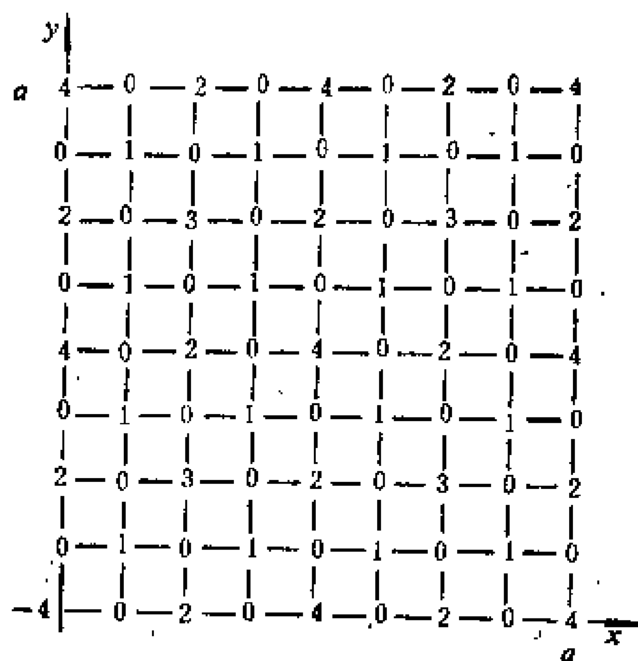


图8.1

设G_k网格层上的局部离散格式记为$[T]_k$，$[T]_0$就是定义在G_0上的5点差分格式，则经相应共轭格式作用后的离散格式在G_k(k = 0, 1, 2, 3)上的表现为

$$[T]_1=[\overline{T}]_0[T]_0=\begin{pmatrix} & & -1 & & \\ & -2 & 0 & -2 & \\ -1 & 0 & 12 & 0 & -1 \\ & -2 & 0 & -2 & \\ & & -1 & & \end{pmatrix}_0=\begin{pmatrix} -1 & -2 & -1 \\ -2 & 12 & -2 \\ -1 & -2 & -1 \end{pmatrix}_1,$$

$$[T]_2=[\overline{T}]_1[T]_1=\begin{pmatrix} -1 & 2 & -1 \\ 2 & 12 & 2 \\ -1 & 2 & -1 \end{pmatrix}_1\begin{pmatrix} -1 & -2 & -1 \\ -2 & 12 & -2 \\ -1 & -2 & -1 \end{pmatrix}_1$$

$$
=\begin{pmatrix}
1 & 0 & -2 & 0 & 1 \\
0 & -32 & 0 & -32 & 0 \\
-2 & 0 & 132 & 0 & -2 \\
0 & -32 & 0 & -32 & 0 \\
1 & 0 & -2 & 0 & 1
\end{pmatrix}_1
$$

$$
=\begin{pmatrix}
 & & 1 & & \\
 & -2 & -32 & -2 & \\
1 & -32 & 132 & -32 & 1 \\
 & -2 & -32 & -2 & \\
 & & 1 & &
\end{pmatrix}_2, \tag{8.11}
$$

$$
[T]_3=[\bar{T}]_2[T]_2=\begin{pmatrix}
1 & -4 & 6 & -4 & 1 \\
-4 & -752 & -2584 & -752 & -4 \\
6 & -2584 & 13348 & -2584 & 6 \\
-4 & -752 & -2584 & -752 & -4 \\
1 & -4 & 6 & -4 & 1
\end{pmatrix}_3 .
$$

从(8.8)式—(8.11)式不难看出，任一定义在G_k($k=0$，1，2，…)网格上对称的局部离散格式$[T]_k$，在其相应的共轭格式的作用下，是一个定义在G_k上的新的局部离散格式，这新格式中有规则地分布着零元素，若将这个新格式旋转$\pi/4$且去掉零元素，则是一个定义在G_{k+1}上的离散格式$[T]_{k+1}$；另外，$[T]_k$($k>0$)的所有元素都可由$[T]_0$生成；最后还可以看到，任一格式$[T]_k$的相应奇格式$[T]_k^-$，只涉及到G_{k+1}网格上的点（因$G_{k+1}\subset G_k$）而并不涉及G_k/G_{k+1}上的网格点。

2.2 RB分划下的缩减方法

讨论数值求解常系数椭圆型方程第一边值问题

$$
\begin{cases}
Lu(x)=f(x), & x\in\Omega, \\
u(x)=g(x), & x\in\partial\Omega,
\end{cases} \tag{8.12}
$$

其中，$\Omega=[0,\ a]\times[0,\ a]$，$x=(x_1,\ x_2)$。

对定解区域Ω作RB分划。先将$[0,\ a]\times[0,\ a]$均分为$2^q\times$

2^{δ}等份，得 $\Omega_0=\Omega\cap G_0$，$h=a/2^{\delta}$。记 $\Omega_k=\Omega\cap G_k$($k=0, 1, 2, \cdots, m$)，$m=2(\delta-1)$，并记 $R_k=\Omega_k/\Omega_{k+1}$($k=0,1,2,\cdots,m-1$)，那么

$$\Omega_0=\Omega_\mu+\sum_{k=0}^{\mu-1}R_k,\qquad \mu=1,\ 2,\cdots,m. \tag{8.13}$$

具体地说，Ω_m 只含有一个内网格点 $(a/2,\ a/2)$；R_{m-1}有 4 个内节点$(a/4,\ a/4)$、$(a/4,\ 3a/4)$、$(3a/4,\ a/4)$和 $(3a/4,\ 3a/4)$；另外，R_k中内节点的数目是 2^{m-k+1}($k=1,\ 3,\ 5,\cdots,m-1$)。

用有限元方法或差分方法对(8.12)式离散处理，可得Ω_0上的离散方程且用局部离散格式表示为

$$\begin{cases}S_0u(x)=r_0(x), & x\in\Omega_0,\\ u(x)=g(x), & x\in\partial\Omega_0,\end{cases} \tag{8.14}$$

其中，S_0表示定义在 Ω_0 网格层上的离散格式(S_1 的意义类同)；另外要注意的是，这里的 S_0 和 $r_0(x)$ 都是应离散方法而异的。

正如(8.8)式～(8.11)式所示，Ω_k 上的局部离散格式在它相应的共轭格式的作用下可变为一个定义在 Ω_{k+1} 上的离散格式，由此我们可得到一串缩减方程：

$$S_ku(P)=r_k(P),\quad P\in\Omega_k,\quad k=m,\ m-1,\cdots,1,\ 0, \tag{8.15}$$

其中，

$$\begin{cases}S_k=\bar{S}_{k-1}S_{k-1}, & k=m,\ m-1,\cdots,1,\\ r_k(P)=\bar{S}_{k-1}r_{k-1}(P), & k=m,\ m-1,\cdots,1,\ P\in\Omega_k.\end{cases} \tag{8.16}$$

这样，我们就可以把§1中的缩减方法推广到二维情形并用于数值求解(8.14)式。

所谓μ次缩减方法就是求解缩减方程组(E_μ)和(F_k)：

$$S_\mu u(P)=r_\mu(P),\qquad P\in\Omega_\mu, \tag{E_μ}$$

$$S_k^{+}u(P)=S[u(P)+r_k(P),\quad P\in R_k,\quad k=\mu-1,\cdots,2,1,0. \tag{F_k}$$

当 $\mu=m$ 时，称之为**完全缩减方法**。

完全缩减方法的实现过程是这样的：方程 (E_m) 是只有一个未知量 $u(Q)$ ($Q\in\Omega_m$)的显示方程，求出 $u(Q)$ ($Q\in\Omega_m$)后，由于

S_{m-1}^{-} 只涉及 Ω_m 中的点，所以方程(F_{m-1})只是关于未知量 $u(P)$ $(P\in R_{m-1})$的低维方程组，用通常的迭代法就可达到快速求解的目的；若 Ω_k 层上的数值解都已获得，因方程 (F_{k-1}) 中S_{k-1}^{-} 的元素只是关于 Ω_k 中的点非零，故方程(F_{k-1})的右端项$S_{k-1}^{-}u(P)+r_{k-1}(P)$ 是已知的，(F_{k-1}) 是关于 $u(P)(P\in R_{k-1}=\Omega_{k-1}/\Omega_k)$ 的方程组，用迭代法解这个低维方程组；如此继续，直至求出 Ω_0 层上的全部数值解。

μ 次缩减方法的矩阵表示见图 8.2，其中 S_k、S_k^{+} 和 S_k^{-}是相

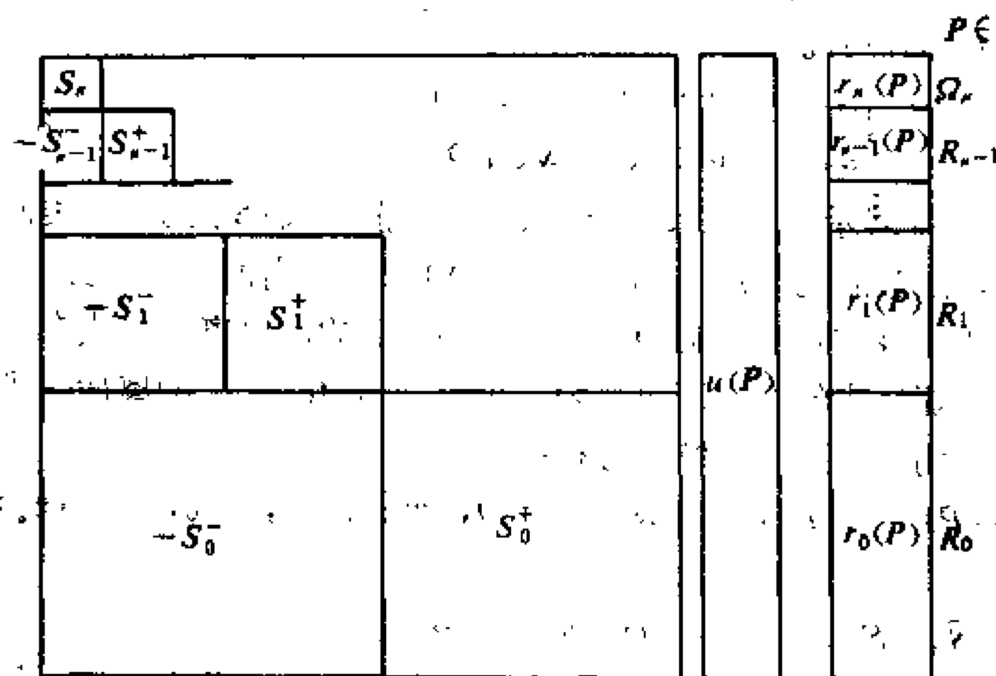


图 8.2

应离散格式在离散方程系数矩阵中的表现。

下面用一简例统一说明在缩减方法中所用到的不同网格层、各种离散格式和缩减方法的矩阵结构。设将一正方形分划为4×4等份，按 RB 规则得网格层序列 Ω_0、Ω_1 和 Ω_2，参见图 8.3。同时得知 $\Omega_i(i=0,\ 1,\ 2)$上的离散格式

$$S_i u(P)=r_i(P)$$

的有关形式如下：



图8.3

$$S_0^- = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}, \quad S_0^+ = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad S_1^- = \begin{pmatrix} 0 & 2 & 0 \\ 2 & 0 & 2 \\ 0 & 2 & 0 \end{pmatrix},$$

$$S_1^+ = \begin{pmatrix} -1 & 0 & -1 \\ 0 & 12 & 0 \\ -1 & 0 & -1 \end{pmatrix}, \quad S_2 = \begin{pmatrix} & & 1 & & \\ & -2 & -32 & -2 & \\ 1 & -32 & 132 & -32 & 1 \\ & -2 & -32 & -2 & \\ & & 1 & & \end{pmatrix},$$

$$r_0(P) = h^2 f(P), \quad r_1(P) = \bar{S}_0 r_0(P), \quad r_2(P) = \bar{S}_1 r_1(P).$$

利用上述表示式可得到3个缩减方程(E_0)、(E_1)和(E_2)，这样便可得到形如图8.2的线性方程组：

$$\begin{pmatrix} 128 & & & & & & & & \\ -2 & 14 & -1 & -1 & & & & & \\ -2 & -1 & 14 & & -1 & & & & \\ -2 & -1 & & 14 & -1 & & & & \\ -2 & & -1 & -1 & 14 & & & & \\ -1 & -1 & -1 & & & 4 & & & \\ -1 & -1 & & -1 & & & 4 & & \\ -1 & & -1 & & -1 & & & 4 & \\ -1 & & & -1 & -1 & & & & 4 \end{pmatrix} \begin{pmatrix} u_{3,3} \\ u_{2,2} \\ u_{2,4} \\ u_{4,2} \\ u_{4,4} \\ u_{2,3} \\ u_{3,2} \\ u_{3,4} \\ u_{4,3} \end{pmatrix} = \begin{pmatrix} r_2(P_{3,3}) \\ r_1(P_{2,2}) \\ r_1(P_{2,4}) \\ r_1(P_{4,2}) \\ r_1(P_{4,4}) \\ r_0(P_{2,3}) \\ r_0(P_{3,2}) \\ r_0(P_{3,4}) \\ r_0(P_{4,3}) \end{pmatrix} \begin{matrix} \}\Omega_2 \\ \\ \\ R_1 \\ \\ \\ \\ R_0 \\ \\ \end{matrix}$$

注意，这个方程组系数矩阵前面5个对角元素与局部离散格式S_2和S_1的中心元素有所不同，这是由于在使用局部离散格式时，要用到界外节点值而利用了反对称假设的缘故(具体参见图8.4及其说明)。这个缩减方程组是很容易求解的。首先由$S_2u(P)=r_2(P)$，$P\in\Omega_2$显式计算出$u_{1,2}$；再考虑解$S_1u(P)=r_1(P)$，$P\in\Omega_1$，由于S_1^-只涉及边界点和Ω_2中的点（这些点处的数值解已知)，故只需求解方程组$S_1^+u(P)=S_1^-u(P)+r_1(P)$，$P\in R_1$；最后，由于同样的理由解方程组$S_0^+u(P)=S_0^-u(P)+r_0(P)$，$P\in R_0$，但这是非常简单的显式计算。

总之，缩减方法的思想在于以下两个方面。第一，利用网格点集的直和形式$\Omega_0=\Omega_\mu+\sum\limits_{k=0}^{\mu-1}R_k$，将复杂的数值求解问题(在$\Omega_0$上)转化为若干个低维集合($R_k$，$k=\mu-1,\cdots,1,0$)上的数值求解问题，这是分裂算法思想的一种应用。第二，为了实现在低维集合上数值求解(F_k)方程，就必须保证Ω_k层上的数值解具有与Ω_0层上数值解相同的精度，缩减方程满足这个要求。

2.3 缩减方法的实施

要实施缩减方法需做下面两部分工作。

一、产生并存贮S_k和$r_k(P)$， $P\in R_k$，$k=0,1,\cdots,m$，(这里简记$\Omega_m=R_m$)。

选定一种离散化方式(差分方法或有限元方法)，就有最细网层上的离散化方程$S_0u(P)=r_0(P)$。再按

$S_k=\bar{S}_{k-1}S_{k-1}$，$r_k(P)=\bar{S}_{k-1}r_{k-1}(P)$，　$P\in R_k$，$k=1,2,\cdots,m$，计算并存贮S_k的有效部分和r_k，计算并存贮这些信息是可用较少的工作量快速且独立地完成的。用多维数组或一维数组存贮S_k和r_k。有了这些数据的计算和存贮，就为缩减方法中逐个地求解(E_m)、(F_{m-1})、…、(F_1)和(F_0)准备了条件。

S_k的有效存贮问题。读者不难发现，S_k的元素分布(这里仅以S_0是5点差分格式为例，其它离散格式类同)有如下几个特点：

(1) S_k中非零元素的个数随着k的增加而急剧增加，其增长规律是

当k为偶数时，S_k的非零元素总数是$2^{k+1}+2^{k/2+1}+1$；

当k为奇数时，S_k的非零元素总数是$2^{k+1}+2^{(k+3)/2}+1$。

(2) S_k中非零元素的量级有巨大差别，其最大元素$s_{0,0}>10^t(t=2^{k-1})$，最小的则是10^0量级。

(3) S_k中非零元素的绝对值$|s_{ij}|$随着$|i|$和$|j|$的增加而急剧减少。例如对于S_{10}就有

$$\sum_{|i|+|j|>10}|s_{ij}|<10^{-8}s_{0,0}.$$

(1)和(2)表明，在缩减方法的实施中，若完整地使用S_k，则会给计算带来极大的不方便；(2)和(3)表明，当k较大时，可使用简化的局部离散格式而不用完整的S_k，这将会大大简化缩减方法的实施过程。为此，可设置一个简化因子σ，当$|s_{ij}|<\sigma s_{0,0}$时就令$s_{ij}=0$；当然，这种简化了的局部离散格式既要保证S_k的基本特征，又要保证以此格式计算的数值解精度。当采用单字长浮点运算时，由于机器的舍入误差是10^{-8}左右，所以可取$\sigma=10^{-8}$；或将$|i|>3$，$|j|>3$，$|i|+|j|>4$的那些s_{ij}忽略，也是可取的方案之一(这相当于取$\sigma=10^{-8}$的情形)。另外，$s_{0,0}$增长过大会超出存贮字长，所以可用一个参数$\left(例如\frac{1}{16}\right)$乘细网层离散格式，以保证$S_k$的元素在$10^{-8}\sim10^{23}$之间。

二、缩减方程的迭代求解

$S_m u(P)=r_m(P)$，$P\in\Omega_m$只是求解一个未知量的方程。经若干步计算，若$u(P)$ $(P\in\Omega_1)$已全部算出，则$S_0u(P)=r_0(P)$，$P\in R_0$，也只是全部由显式方程构成的方程组，它们的求解是极容易的。

对于方程(F_{m-1})，…，(F_1)，可采用任一种逐点迭代法求解，例如GS方法，求解过程是快速收敛的，这是因为方程组(F_k) $(k=m-1,\cdots,2,1)$的系数矩阵S_k^+是强对角占优的缘故，

当然，读者可用局部Fourier 分析法求得迭代收敛因子。§4.2中的分析表明，求解$(F)_k$，$k=m-1$，…，1 的迭代过程是 h 无关收敛的，且只要8～9步迭代就够了。

下面讨论完全缩减方法求解二维椭圆边值问题所需的计算量。

设 S_k 中共有Q个元素，Ω_0 共有N个节点，按 RB 分划共有 m个网格层；则 R_k 中最多有 $N/2^k$ 个节点，计算(F_k)的右端项 $r_k=\bar{S}_{k-1}r_{k-1}$至多需用计算量 $QN/2^k$。因求解 (F_k) 的迭代步对一切 k 有上界 $l(\varepsilon)$ (参见 §4.2)，则求解 (F_k) 所需的工作量不会超过 $l(\varepsilon)QN/2^k$。故采用缩减方法求解(6.14)式的总工作量为

$$\sum_{k=0}^{m}(1+l_k(\varepsilon))QN/2^k\leqslant(1+l(\varepsilon))QN\sum_{k=0}^{m}\frac{1}{2^k}=O(N)。$$

如果 S_0 是 5 点离散格式，则总工作量的精细估计如下：形成缩减方程的总工作量不超过$9\frac{1}{4}N$个乘法运算和$9\frac{1}{4}N$个加法运算，求解缩减方程的总工作量不超过$5\frac{7}{8}N$个除法、$5\frac{7}{8}N$个乘法和$22\frac{5}{8}N$个加法运算。

下面给出几个用缩减法求解的数值例子：

$$\begin{cases}-\Delta u=f, & (x,\ y)\in\Omega,\\ u=g, & (x,\ y)\in\partial\Omega,\end{cases}$$

其中 $\Omega=(0,\ 1)\times(0,\ 1)$。对于情形：

(1) $f=0$，$g=1$；

(2) $f=0$，$g=\frac{1}{4}[x(1-x)+y(1-y)]$；

(3) $f=1$，$g=0$；

(4) $f=\cos(100xy)$，$g=x+y$；

(5) $f=\exp(10e^x)\sin(10y)$，$g=xy$；

用缩减方法求解的数值结果列于表8.2中；表中的 n 表示[0,1]被

均分的等份数，相对误差 $\delta_n=\|u'-u\|_\infty/\|u\|_\infty$；表8.2中的数据说明缩减方法是数值稳定的。

表8.2　用缩减法对不同数例的计算结果比较

n \ δ_n \ (a)	(1)	(2)	(3)	(4)	(5)
8	0.24E−5	0.19E−5	0.10E−5	0.29E−5	0.50E−6
16	0.16E−4	0.13E−4	0.10E−4	0.10E−4	0.22E−5
32	0.26E−4	0.22E−4	0.21E−4	0.18E−4	0.57E−5
64	0.38E−4	0.31E−4	0.29E−4	0.24E−4	0.13E−4
128	0.43E−4	0.39E−4	0.40E−4	0.31E−4	0.23E−4
256	0.52E−4	0.45E−4	0.48E−4	0.37E−4	0.32E−4

最后，我们讨论缩减方法的应用问题。

对于矩形区域矩形分划情形，无论是固定边值条件的椭圆型方程还是周期边值条件的椭圆型方程，应用缩减方法数值求解是毫无困难的。但此时应该注意，缩减方程的离散算子 S_k^+ 和 S_k^- 以及其右端项 r_k 的元素可能会涉及到定解区域Ω外的节点及其函数值，此时可假设网格可以向四周延拓，且假设椭圆边值问题的解也能周期延拓，这样缩减方法就可以没有困难地实行了。例如，对于第一边值条件，此时可设想解函数作双周期延拓，即可作反对称假设(参见图8.4)：

$$u(P)=2u(Q)-u(\hat{P}),$$

$$u(P')=2u(Q')-u(\hat{P}).$$

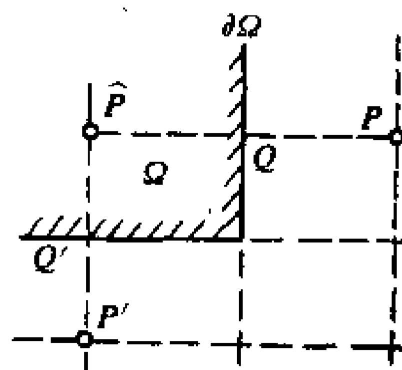


图8.4

又例如，$\bar{\Omega}=[0,\ 1]\times[0,\ 1]$，微分方程右端项是周期函数且具有周期边界条件：

$$u(0,\ y)=u(1,\ y),$$

$$u(x,\ 0)=u(x,\ 1),$$

$$u_x(0,\ y)=u_x(1,\ y),\qquad u_y(x,\ 0)=u_y(x,\ 1);$$

此时则可按微分方程解的周期特点延拓。

对于一般的曲边定解区域，应用该方法尚有一定困难。但若在矩形网格分划下将曲边定解区域作适当修改，使其在边界上与矩形分划相适应，则缩减方法还是可以应用的。此时对要用到边界外的节点函数值仍可作反对称假设，参见图8.4。

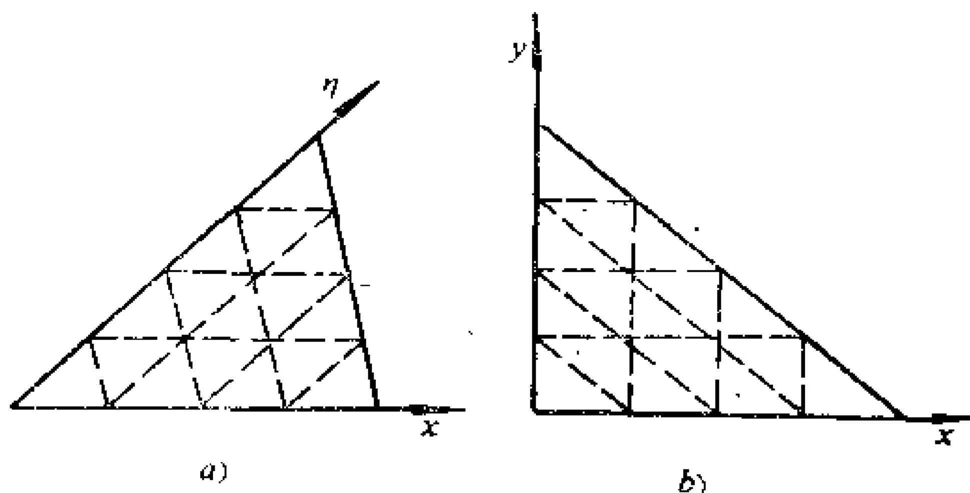


图8.5

对于由平行网线构成的三角分划情形，如图8.5a)所示，只要利用一个简单的仿射变换，即有如图8.5b)所示的情形。所以，格式的书写形式仍可按 $x-y$ 坐标系下节点的排列规则写成；计算时，可将图8.5a)中的节点按 x 和 η 方向编号；此时，网格层序列$\{\Omega_l\}$的构造仍相同于正方形网格RB分划的情形。

§3 单向缩减方法

设 G 是如同(8.4)式定义的 R^2 的无限网格，$P_0=(x_0,\ y_0)$，$P_{ij}=(ih,\ jh')$，且记

$$\begin{cases}G_x:=\{x\in R\,|\,x=x_0+ih;\ i=0,\ \pm1,\ \pm2,\ \cdots\},\\ G_y:=\{y\in R\,|\,y=y_0+jh';\ j=0,\ \pm1,\ \pm2,\ \cdots\},\end{cases}\tag{8.17}$$

则有 $G=G_x\times G_y$。

在 G 上可构造网格层序列$\{G_k\}$，k = 0，1，⋯，其中

$$G_k = G_{x,k} \times G_y, \tag{8.18}$$

$$G_{x,k} = \{x \in R \mid x = x_0 + ih_k;\ h_k = 2^k h;\ i = 0, \pm 1, \pm 2, \cdots\};$$

也就是说

$$G_k = \{P \in R^2 \mid P = P_0 + P^k_{ij};\ P^k_{ij} = (ih_k,\ jh');\ \},$$
$$i,\ j = 0,\ \pm 1,\ \pm 2,\ \cdots$$

此时相当于(8.4)式中的 $P_{ij} = P^k_{ij} = (ih_k,\ jh')$.在这种理解下，$G_k$ 上局部离散算子 S_k 的定义(8.5)式和(8.6)式仍然成立。

定义在G_k上的局部离散算子 S_k 可以按行分解为"行"离散算子 X_j(定义在G_x上)的和：

$$S_k = \sum_j E_j X_j = \cdots + E_{-1}X_{-1} + E_0 X_0 + E_1 X_1 + \cdots, \tag{8.19}$$

其中，E_j 是与 y 有关的离散算子，且 $E_j v(x,\ y) = v(x,\ y + jh')$。$S_k$ 的**x-共轭算子**$\bar{S}_k$、**x-偶算子**S_k^+和**x-奇算子**S_k^-同样可定义为

$$\bar{S}_k = \sum_j E_j \bar{X}_j,\quad S_k^+ = \sum_j E_j X_j^+,\quad S_k^- = \sum_j E_j X_j^-, \tag{8.20}$$

其中，$\bar{X}_j$、X_j^+ 和 X_j^- 分别是一维局部离散算子 X_j 所对应的共轭、偶和奇的局部离散算子。

例如，5点离散格式在G_k上可表示为

$$S_k = E_{-1}X_{-1} + E_0 X_0 + E_1 X_1$$
$$= \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} [0\ \ -1\ \ 0]_k + \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} [-1\ \ 4\ \ -1]_k + \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} [0\ \ -1\ \ 0]_k;$$

$$\bar{S}_k = \begin{pmatrix} & -1 & \\ 1 & 4 & 1 \\ & -1 & \end{pmatrix}_k,\quad S_k^+ = \begin{pmatrix} & -1 & \\ 0 & 4 & 0 \\ & -1 & \end{pmatrix}_k,\quad S_k^- = \begin{pmatrix} & 0 & \\ 1 & 0 & 1 \\ & 0 & \end{pmatrix}_k;$$

$$\bar{S}_k S_k = \begin{pmatrix} & & 1 & & \\ & & -8 & & \\ -1 & 0 & 16 & 0 & -1 \\ & & -8 & & \\ & & 1 & & \end{pmatrix}_k = \begin{pmatrix} & 1 & \\ & -8 & \\ -1 & 16 & -1 \\ & -8 & \\ & 1 & \end{pmatrix}_{k+1}.$$

又如，9点差分格式在G_k上可表示为

$$S_k=\begin{pmatrix}-1 & -4 & -1\\ -4 & 20 & -4\\ -1 & -4 & -1\end{pmatrix}_k;\quad \bar{S}_k=\begin{pmatrix}1 & -4 & 1\\ 4 & 20 & 4\\ 1 & -4 & 1\end{pmatrix}_k;$$

$$\bar{S}_kS_k=\begin{pmatrix}-1 & 0 & 14 & 0 & -1\\ -8 & 0 & -176 & 0 & -8\\ 18 & 0 & 396 & 0 & 18\\ -8 & 0 & -176 & 0 & -8\\ -1 & 0 & 14 & 0 & -1\end{pmatrix}_k=\begin{pmatrix}-1 & 14 & -1\\ -8 & -176 & -8\\ 18 & 396 & 18\\ -8 & -176 & -8\\ -1 & 14 & -1\end{pmatrix}_{k+1}.$$

由这两个例子可见，$\bar{S}_k\cdot S_k$是一个定义在G_k上的局部离散算子。正由于此，我们可以构造数值求解离散齐次边值问题

$$\begin{cases}S_0u(P)=r_0(P), & P\in\Omega_0,\\ u(P)=g(P), & P\in\partial\Omega_0\end{cases}$$

的单向缩减方法，其中$\Omega_0=\Omega\cap G_0$，$\partial\Omega_0=\partial\Omega\cap G_0$，$\Omega=\{(x,y)\mid 0<x<a,\ 0<y<b\}$，$G_0=G_x\times G_y$，$x_0=y_0=0$，$h=a/n$，$h'=b/n$，$n=2^m$；另外还假设定义在$G_x\cap(0,\ 1)\times G_y$上的网格函数$u$以及$r_0$在$y=0$及$y=b$处是反对称的。

利用§2中的处理办法，可推导出x方向缩减方程：

$$S_ku(P)=r_k(P),\quad P\in\Omega_k,\ k=m-1,\ \cdots,\ 1,\ 0,\qquad(8.21)$$

其中

$$S_k=\bar{S}_{k-1}S_{k-1},\quad k=m-1,\ \cdots,\ 2,\ 1,$$

$$r_k(P)=\bar{S}_{k-1}r_{k-1}(P),\ P\in\Omega_k,\ k=m-1,\cdots,\ 2,\ 1.$$

例如，对于5点离散格式情形，有

$$S_ku(P)=-u(x-h_k,\ y)+C_ku(x,\ y)-u(x+h_k,\ y)$$

其中，C_k是定义在G_k上的离散算子(加方框记号的是中心元素)：

$$C_0=[-1\ \ \boxed{4}\ \ -1],\ C_1=[1\ \ -8\ \ \boxed{16}\ \ -8\ \ 1]^T,$$

$$C_2=[1\ \ -16\ \ 96\ \ -272\ \ \boxed{384}\ \ \text{对称}]^T.$$

又如，对于9点差分格式情形，有

$$S_k u(P) = -B_k u(x-h_k, y) + C_k u(x, y) - B_k u(x+h_k, y),$$

其中

$C_0 = [-4 \quad \boxed{20} \quad -4]^T$, $\qquad B_0 = [-1 \quad \boxed{-4} \quad -1]^T$,

$C_1 = [14 \quad -176 \quad \boxed{396}$ 对称$]^T$, $\quad B_1 = [-1 \quad -8 \quad \boxed{-18}$对称$]^T$,

$C_2 = [194 \quad -4960 \quad 41864 \quad -144928 \quad \boxed{218252}$ 对称$]^T$,

$B_2 = [-1 \quad -16 \quad -100 \quad -304 \boxed{-454}$ 对称$]$。

最后，令 $\Omega_k = \Omega \cap G_k (k=0, 1, \cdots, m-1)$，$R_k = \Omega_k / \Omega_{k+1}$ $(k=0, 1, \cdots, m-1)$，G_k如(8.18)式定义，则可将(8.21)式进一步分解，得到便于计算的μ次x方向缩减方程$(0 \leqslant \mu \leqslant m-1)$：

$$\begin{cases} S_\mu u(P) = r_\mu(P), & P \in \Omega_\mu, \quad (E_\mu) \\ S_k^+ u(P) = S_k^- u(P) + r_k(P), & P \in R_k. \quad (F_k) \end{cases}$$

当 $\mu = m-1$ 时，上述方法称为完全的 x 方向缩减方法。此时由方程组(E_μ)解出数值解$u\left(\frac{a}{2}, jh'\right)$，$j=1, 2, \cdots, n-1$；由方程组$(F_{\mu-1})$解出数值解$u\left(\frac{a}{4}, jh'\right)$和$u\left(\frac{3a}{4}, jh'\right)$，$j=1, 2, \cdots, n-1$；这样继续下去，可逐“列”求得数值解。

综上所述，单向缩减方法的构造和实施，基本上类似于§2中所述的缩减方法；当然，§2中所讨论的关于缩减方法的应用范围，例如齐次边值条件和周期边值条件等情形，对于单向缩减方法也是基本适用的。

§4 误差估计

4.1 完全缩减方法的相对误差估计

正如前述，缩减方法在实施中产生误差的原因主要是由于采用了简化格式 $S_k(\sigma)$ 以及机器精度和舍入误差存在的缘故。为了作误差估计，先给出一些有关的模量定义和记号。因为局部离散

格式 S_k 对应着相应离散方程组的系数矩阵 $L_k=(a_{\beta,\gamma})$，所以 $\|L_k\|_\infty=\max\limits_{\beta}\sum\limits_{\gamma}|a_{\beta,\gamma}|$ 可用 $\|S_k\|_\infty$ 表示且简记为 $\|S_k\|$。相应地，$\|S_k(\sigma)\|$ 和 $\|(S_k^+)^{-1}\|$ 等模量也有类似解释。另外还约定 $\|v_k\|^{\Omega(j)}$ 表示 $v_k\in\Omega_k$ 限制在 Ω_j 上的矢量模，若 $j=k$，就简记为 $\|v_k\|$。还记 $|S|=\sum\limits_{i,j}|s_{ij}|$。

假设缩减方程左端是准确的，右端存在误差，即

$$S_k\tilde{u}(P)=\tilde{r}_k(P),\qquad k=0,\ 1,\ \cdots,\ m; \tag{8.22}$$

假设缩减方程右端项误差是由于 r_0 受到小扰动 δr_0 影响后，又受到缩减过程中的简化格式和舍入误差的影响而产生的；则 $\tilde{r}_k(P)$ 适合如下关系：

$$\tilde{r}_0=r_0-\delta r_0,$$

$$\tilde{r}_k=G\cdot\bar{S}_{k-1}(\sigma)\tilde{r}_{k-1}(P),\qquad P\in\Omega_k,\ k=1,\cdots,m,$$

其中 G 表示相应于舍入影响的离散算子。

若记 $\Delta r_k=r_k-\tilde{r}_k$，$\Delta\bar{S}_k=S_k-G\cdot\bar{S}_k(\sigma)$，$\delta r_k=\Delta\bar{S}_{k-1}\tilde{r}_{k-1}$，则误差函数 Δr_k 和 $\delta r_k(k=1,\ \cdots,\ m)$ 适合下述关系：

$$\begin{aligned}\Delta r_k=r_k-\tilde{r}_k&=\bar{S}_{k-1}r_{k-1}-G\cdot\bar{S}_{k-1}(\sigma)\tilde{r}_{k-1}\\&=\bar{S}_{k-1}r_{k-1}-\bar{S}_{k-1}\tilde{r}_{k-1}+\bar{S}_{k-1}\tilde{r}_{k-1}\\&\quad-G\cdot\bar{S}_{k-1}(\sigma)\tilde{r}_{k-1}\\&=\bar{S}_{k-1}\Delta r_{k-1}+\Delta\bar{S}_{k-1}(r_{k-1}-\Delta r_{k-1}),\end{aligned} \tag{8.23}$$

$$\delta r_{k-1}=\Delta\bar{S}_{k-1}\tilde{r}_{k-1}=\Delta\bar{S}_{k-1}(r_{k-1}-\Delta r_{k-1}). \tag{8.24}$$

在只考虑缩减过程误差的情况下，缩减方程 (8.22) 是在 R_k $(k=m,\ \cdots,\ 1,\ 0)$ 上顺次单独地被解出的，所以若记 $\delta_k u=u-\tilde{u}_k$，即知 $\delta_k u$ 是由方程组

$$\begin{cases}S_j\tilde{u}(P)=\tilde{r}_j(P), & j=m,\ \cdots,\ k+1,\\ S_k\delta_k u=\delta r_k, & \\ S_j u(P)=0, & j=k-1,\ \cdots,\ 0\end{cases} \tag{8.25}$$

决定的。总之，可得到缩减误差方程和缩减过程的总误差为

$$S_k\delta_k u=\delta r_k,\qquad k=0,\ 1,\ \cdots,\ m, \tag{8.26}$$

$$\delta u = \delta_0 u + \delta_1 u + \cdots + \delta_m u. \tag{8.27}$$

下面讨论$\|\delta_k u\|^{\Omega(k)}$和$\|\delta_k u\|^{\Omega(0)}$的关系。假定

$$\|(\Delta S)v\| \leqslant \eta |S| \|v\|, \quad \|\delta r_0\| = \|\Delta r_0\| \leqslant \varepsilon \|r_0\| \quad (\varepsilon \geqslant 0);$$

并利用(8.23)式和(8.24)式，有下面不等式对一切k = 1, …, m成立

$$\begin{cases} \|\Delta r_k\| \leqslant |\bar{S}_{k-1}| (\eta \|r_{k-1}\| + (1+\eta) \|\Delta r_{k-1}\|), \\ \|\delta r_k\| \leqslant \eta |\bar{S}_{k-1}| (\|r_{k-1}\| + \|\Delta r_{k-1}\|); \end{cases}$$

利用上述不等式和$S_j u = r_j$，可进一步估计

$$\|r_j\| \leqslant |S_j| \|u\|^{\Omega(j)} \leqslant |S_j| \|u\|^{\Omega(0)};$$

若注意递推公式

$$\begin{cases} \Delta_0 = \varepsilon a_0, \quad \Delta_k = \eta q_{k-1} a_{k-1} + (1+\eta) q_{k-1} \Delta_{k-1} \\ \qquad\qquad\qquad\qquad\qquad\qquad (k = 1, \cdots, m-1), \\ \delta_0 = \varepsilon a_0, \quad \delta_k = \eta q_{k-1}(a_{k-1} + \Delta_{k-1}) \quad (k = 1, \cdots, m), \end{cases} \tag{8.28}$$

其中

$$a_j = \|S_j^{-1}\| |S_j|, \qquad q_j = |\bar{S}_j| \|S_{j+1}^{-1}\| / \|S_j^{-1}\|,$$

就可归纳得到$\|S_k^{-1}\| \|\delta r_k\| \leqslant \delta_k \|u\|^{\Omega(0)} (k = 0, \cdots, m)$；结合(8.26)式，有

$$\|\delta_k u\|^{\Omega(k)} \leqslant \delta_k \|u\|^{\Omega(0)}. \tag{8.29}$$

为了进一步估计$\|\delta_k u\|^{\Omega(0)}$(k≥1)，由(8.25)式知

$$S_j^{+} \delta_j u = S_j^{-} \delta_j u, \qquad j = 0, 1, \cdots, k-1,$$

注意$\delta_j u$(j = 0, 1, …, k − 1)是由$\delta_k u$不变地传播得到的，所以有$S_j^{+} \delta_k u = S_j^{-} \delta_k u$ $\quad \|\delta_k u\|^{R(k)} \leqslant \|(S_j^{+})^{-1}\| |S_j^{-}| \|\delta_k u\|^{\Omega(j+1)} \quad P \in R^k$，$j$ = 0, 1, …, k − 1，从而有

$$\|\delta_k u\|^{\Omega(j)} \leqslant \kappa_j \|\delta_k u\|^{\Omega(j+1)}; \qquad j = 0, 1, \cdots, k-1,$$

再经简单递推，就有

$$\|\delta_k u\|^{\Omega(0)} \leqslant \xi_{k-1} \|\delta_k u\|^{\Omega(k)}, \tag{8.30}$$

其中

$$\xi_j = \kappa_j \cdot \kappa_{j-1} \cdots \kappa_0, \quad \kappa_j = \max\{\|(S_j^{+})^{-1}\| |S_j^{-}|, 1\}. \tag{8.31}$$

综合(8.29)式和(8.30)式，有

$$\|\delta_k u\|^{\Omega(0)}/\|u\|^{\Omega(0)}\leqslant\xi_{k-1}\delta_k. \qquad (8.31)$$

最后利用(8.27)和(8.31)式得关于完全缩减法相对误差估计定理。

定理8.1 在缩减过程中，由于简化离散格式和机器运算精度原所产生的缩减总误差为

$$\|\delta u\|^{\Omega(0)}/\|u\|^{\Omega(0)}\leqslant\delta_0+\xi_0\delta_1+\cdots+\xi_{m-1}\delta_m, \qquad (8.32)$$

其中，δ_j 由(8.28)式确定，ξ_j 由(8.31)式确定。

注[40]：(8.32)式可改写为

$$\|\delta u\|^{\Omega(0)}/\|u\|^{\Omega(0)}\leqslant\tilde{L}+\tilde{Q},$$

其中，$\tilde{L}=\varepsilon\alpha_0+\eta\sum_{k=0}^{m-1}\xi_k q_k\alpha_k$，$\tilde{Q}=\eta\sum_{k=0}^{m-1}\xi_k q_k\Delta_k$。$\tilde{L}$ 是关于 ε 和 η 线性的，$\tilde{Q}$ 是非线性的，且 $\tilde{Q}\ll\tilde{L}$。人们可具体估计出 $\tilde{L}\leqslant(0.59\varepsilon+1.37\eta)N$，$\eta\leqslant10^{-4}$，$\varepsilon$ 是机器精度，这样便可估计出整个缩减过程所产生的误差 $\|\delta u\|^{\Omega(0)}/\|u\|^{\Omega(0)}\approx1.4\times10^{-4}N$。

4.2 (F_k)的迭代控制

用某种迭代法求解 (F_k) 可表为

$$\begin{cases}u_k^{i+1}=B_k u_k^i+w_k, & i=1,\ 2,\ \cdots,\\ u_k^0=0,\end{cases}$$

其中 B_k 是相应的迭代矩阵。记 $\beta_k=\rho(B_k)$，则经 i 次迭代后，有

$$\|u_k-u_k^i\|\leqslant\beta_k^i(1-\beta_k)^{-1}\|u_k^1\|.$$

令 $i=1$，得

$$(1-\beta_k)\|u_k-u_k^1\|\leqslant\beta_k^1(\|u_k-u_k^1\|+\|u_k\|).$$

综合这两个式子，即知第 $l(k)$ 次迭代近似解的误差估计

$$\|u_k-u_k^{l(k)}\|R(k)/\|u_k\|R(k)\leqslant\beta_k^{l(k)}(1-2\beta_k)^{-1} \qquad (2\beta<1).$$

由(8.30)式，有

$$\|u_k-u_k^{l(k)}\|^{\Omega(0)}/\|u_k\|^{\Omega(0)}\leqslant\xi_{k-1}\beta_k^{l(k)}(1-2\beta_k)^{-1}\quad(2\beta<1). \qquad (8.33)$$

人们利用 $\|B_k\|_\infty(\geqslant\rho(B_k))$ 估计出

$$0.048 \leqslant \beta_k \leqslant 0.2 \quad (1 \leqslant k \leqslant 8), \quad \beta_k \leqslant 0.1 (9 \leqslant k \leqslant 20). \tag{8.34}$$

(8.34)式表明了β_k的h无关有界性，迭代求解（F_k）的敛速与h无关。(8.33)式可用以确定求解（F_k）所需的迭代步数$l(k)$。例如，取$\sigma = 10^{-8}$，则$\| u_k - u_k^{l(k)} \|^{\Omega(0)} / \| u_k \|^{\Omega(0)} \approx 10^{-8}$，估计出$l_1 = 7$，$l_2 = 4$，$l_3 = 6$，$l_k = 5 (k \geqslant 4)$。

作为本章结束我们指出：缩减方法将细网格层上的数值求解问题转化为粗网格层上的数值求解问题，它也有效地利用了多个网格层上的数值信息，从而取得了高效率效果。该方法利用细网格层上的离散方程和相应的共轭离散格式作恒等变换，这既使数值求解问题顺利地转换到较粗的网格层上，同时又保证了数值解精度在变换后仍满足细网层上数值解精度的要求。这是有启发性的，用这种办法可以构造出高精度离散格式。

特别地，缩减方法、余量校正迭代方法、分裂方法、多层网格方法以及嵌套式迭代方法等，都可以有机地结合起来，以进一步提高实际计算效率和适应性。

例如，在标准粗化网格序列$\{\Omega_i\}$上定义的离散方程

$$L_i u_i = f_i, \qquad (x, y) \in \Omega_i, \quad i = 1, 2, \cdots, l \tag{8.35}$$

受到相应的共轭离散算子作用后，我们就有缩减方程

$$\tilde{L}_{i-1} \tilde{u}_{i-1} = \tilde{f}_{i-1}, \qquad (x, y) \in \Omega_{i-1}, \quad i = 2, 3, \cdots, l. \tag{8.36}$$

显然，与离散方程

$$L_{i-1} u_{i-1} = f_{i-1} \tag{8.37}$$

相比，$\tilde{u}_{i-1}(\Omega_{i-1})$比$u_{i-1}(\Omega_{i-1})$具有更高精度，且

$$\| (u - \tilde{u}_{i-1})(\Omega_{i-1}) \| = \| (u - u_i)(\Omega_{i-1}) \|, \quad i = 2, 3, \cdots, l.$$

这样，在Ω_{i-1}网格层上就有了两个离散精度不相同的离散方程(8.36)和(8.37)式，于是，余量校正迭代原则就可以应用了。这时，用满足(8.35)式精度阶要求的数值解作为初值，记为u_{i-1}^0，作余量校正迭代(参见第七章)：

$$\begin{cases} u_{l-1}^{k+1} = u_{l-1}^{k} - v_{l-1}^{k}, & k = 0, 1, \cdots. \\ L_{l-1} v_{l-1}^{k} = \tilde{L}_{l-1} u_{l-1}^{k} - \tilde{f}_{l-1}, \end{cases} \tag{8.38}$$

当然，(8.38)式可用分裂算法、MG 方法等其它高效率算法求解；经有限步余量校正迭代步后，有

$$\|u - u_{l-1}^{k}\| \approx \|u - \tilde{u}_{l-1}\|.$$

因为 $\tilde{u}_{l-1}$ 和 u_l 的精度阶是相同的，设为 κ 阶，所以

$$\|u - \tilde{u}_{l-1}\| / \|u - u_{l-1}\| \approx 1/2^{\kappa},$$

这就是说，只要经一个余量校正迭代步就行了，这时

$$\|u - u_{l-1}^{1}\| \approx \|u - \tilde{u}_{l-1}\|.$$

总之，对若干有特色的好算法进行分析比较，有机结合；努力发扬优点，克服缺点，就一定会产生立杆见影的高效率数值计算效果。

习　题

1. 利用 Hermite 方法可以构造 $\Delta u = f$ 的高精度差分格式。办法如下：

先作线性组合

$$R = \sum_{(s,t)\in G} [a_{ij} u(P_{ij}) + b_{ij}(\Delta u)_{P_{ij}} + c_{ij}(\Delta^2 u)_{P_{ij}}],$$

其中 a_{ij}、b_{ij}、c_{ij} 是待定系数，$G = \{(i, j\pm 1), (i\pm 1, j), (i\pm 1, j\pm 1)\}$；再将上式各项都在 P_{ij} 处作 Taylor 展开 $(h_1 = h_2 = h)$；整理合并 $h^s (s = 0, 1, 2, \cdots)$ 的同次幂项；令 h^s 的(那些与待定系数有关的）系数全部为零，就得到关于待定系数的方程组；解这个方程组，确定待定系数，有

$$\begin{aligned} R = {} & 20u(P_{ij}) - 4\sum_{(m,n)\in G_1} u(P_{mn}) - \sum_{(m,n)\in G_2} u(P_{mn}) \\ & + \frac{h^2}{5}\Big(34\Delta u(P_{ij}) - \sum_{(m,n)\in G_1} \Delta u(P_{mn})\Big) \\ & + \frac{h^4}{30}\Big(17\Delta^2 u(P_{ij}) + \sum_{(m,n)\in G_1} \Delta^2 u(P_{mn})\Big) = O(h^8), \end{aligned}$$

其中，$G_1=\{(i\pm1,\ j),\ (i,\ j\pm1)\}$，$G_2=\{(i\pm1,\ j\pm1)\}$；最后，两边同除$h^2$，且用$f$和$\Delta f$代替$\Delta u$和$\Delta^2 u$，即可得到$O(h^6)$的差分格式。

试按所述过程，详细推导这个高精度差分格式。并将其表示为局部离散格式。

2. 用子刚阵叠加的办法详细推导矩形双线性元和矩形双二次元的局部离散格式。

3. 推导解 Poisson 方程的相同网格步长 5 点局部离散格式的线性组合

$$4\begin{pmatrix} & -1 & \\ -1 & 4 & -1 \\ & -1 & \end{pmatrix}+\begin{pmatrix} -1 & & -1 \\ & 4 & \\ -1 & & -1 \end{pmatrix}$$

得到的差分格式和截断误差。并就不同网格步长的相应情形讨论之。

4. 用矩形双线性元和 5 点差分格式数值求解 Poisson 方程第一边值问题。

(1) 试编制用 ω-Jacobi 方法迭代求解的子程序；

(2) 试编制用 GS 方法迭代求解的子程序；

(3) 试编制 SOR 方法迭代求解的子程序；

(4) 试编制 RB-GS 方法迭代求解的子程序。

5. 试推导定理 2.2 中的(2.9)式。

6. 试用局部 Fourier 分析法分析 9 点差分格式和正方形双线性元格式的 GS 方法迭代的收敛因子。

7. 试用局部 Fourier 分析法验证：5 点差分格式和 9 点差分格式 SOR 方法的收敛因子都是 $1-O(h)$ 阶的。

8. 对 Poisson 方程的等腰直角三角二次元局部离散格式采用RB-Jacobi 和GS 方法求解。试分别写出对应于单元顶点和单元边中点处的迭代格式，且计算相应的收敛因子。

9. 证明由粗网格校正所决定的迭代过程（参见4.7式）是不收

敛的。

10。在MG方法中，松弛迭代的光滑作用强调对高频分量的衰减使误差变得“光滑”，而不是强调压缩和减小误差，为什么？

11。设U和V是由n维向量构成的向量空间，且定义Euclid内积和Euclid模为

$$\langle x, y\rangle_W = c_W\sum_{i=1}^{n} x_i\bar{y}_i, \qquad \forall x, y\in W,\ W=U\text{或}V,$$

$$\|x\|_W = \left(c_W\sum_{i=1}^{n}|x_i|^2\right)^{1/2}, \qquad \forall x\in W;\ c_W\text{为比例因子,}$$

则$(W, \langle\cdot, \cdot\rangle_W)$是Hilbert空间。若映射$A: U\to V$的伴随映射$A^*: V\to U$定义为

$$\langle Au, v\rangle_V = \langle u, A^*v\rangle_U, \qquad \forall u\in U,\ \forall v\in V,$$

试证明：

$$A^* = (c_V/c_U)\bar{A}^T.$$

12。证明限制算子I_h^{2h}和内插算子I_{2h}^h生成的矩阵

$$[I_h^{2h}] = \frac{1}{4}\begin{pmatrix} 1 & 2 & 1 & & & & \\ & & 1 & 2 & 1 & & \\ & & & \ddots & & & \\ & & & & \ddots & & \\ & & & & 1 & 2 & 1 \end{pmatrix} \text{和} [I_{2h}^h] = \frac{1}{2}\begin{pmatrix} 1 & & & \\ 2 & 1 & & \\ 1 & 2 & & \\ & 1 & \ddots & \\ & & \ddots & 1 \\ & & & 2 \\ & & & 1 \end{pmatrix}$$

是关于内积

$$\langle v_l, w_l\rangle = h_l\sum_{x\in\Omega_l} v_l(x)\overline{w_l(x)}$$

伴随的，即满足关系式$[I_h^{2h}]^* = [I_{2h}^h]$。

13。由第2题知，可简单地用限制算子和内插算子表示由它们生成的矩阵。试证明下列2组限制算子和内插算子是关于内积

$$\langle u_l, v_l\rangle = \sum_{x\in\Omega_l} h_l^2 u_l(x)\overline{v_l(x)}$$

伴随的。

(1) $I_h^{2h}=\frac{1}{16}\begin{pmatrix}1&2&1\\2&4&2\\1&2&1\end{pmatrix}$, $I_{2h}^{h}=\frac{1}{4}\begin{pmatrix}1&2&1\\2&4&2\\1&2&1\end{pmatrix}$.

(2) $I_h^{2h}=\frac{1}{8}\begin{pmatrix}0&1&1\\1&2&1\\1&1&0\end{pmatrix}$, $I_{2h}^{h}=\frac{1}{2}\begin{pmatrix}0&1&1\\1&2&1\\1&1&0\end{pmatrix}$.

14. 叙述第13题(1)中内插算子I_{2h}^{h}的双线性特征。

15. 离散化误差定义为$\tilde{R}_k u-u_k$，其中$\tilde{R}_k$是到网格Ω_k上的某个限制算子。假设某个$\tilde{r}$满足$\tilde{R}_{k-1}=\tilde{r}\tilde{R}_k$，且定义$\tilde{p}$的插值误差为$\tilde{p}\tilde{r}u_k-u_k$。试证明：对于连续问题和相应离散问题的解$u$和$u_k$，若

$\|\tilde{R}_k u-u_k\|_{\Omega_k}=O(h_k^{\alpha})$，$\|\tilde{p}\tilde{r}u_k-u_k\|_{\Omega_k}=O(h_k^{\beta})$，$\|\tilde{p}\|_{\Omega_k}=O(1)$，

则可推导出

$$\|\tilde{p}u_{k-1}-u_k\|_{\Omega_k}\leqslant ch_k^{\min(\alpha,\beta)}.$$

16. 由已知粗网格点$(x,\ y)$，$(x+2h_l,\ y)$，$(x,\ y+2h_l)$和$(x+2h_l,\ y+2h_l)$处的u_{l-1}，用

$$(2h_l^2)^{-1}\begin{pmatrix}-1&0&-1\\0&4&0\\-1&0&-1\end{pmatrix}u_l=f_l$$

计算在$(x+h_l,\ y+h_l)$处的u_l值；在其它网格点处的u_l值用

$$(h_l^2)^{-1}\begin{pmatrix}&-1&\\-1&4&-1\\&-1&\end{pmatrix}u_l=f_l$$

计算之。

(1) 试叙述实现这种内插公式的显式计算过程；

(2) 试证明其插值精度是4阶的。

17. 验证经第13题(1)中插值算子I_{2h}^{h}作用得到的$v_l=I_{2h}^{h}v_{l-1}$与利用v_{l-1}在$(0,\ 0)$，$(2h,\ 0)$，$(0,\ 2h)$，$(2h,\ 2h)$

节点值所作的双线性插值相同；试确定I_{2h}^h的插值精度阶。

18. 验证经第13题(2)中插值算子I_{2h}^h作用得到的$v_l=I_{2h}^h v_{l-1}$与利用v_{l-1}在$(0,0)$，$(0,2h)$，$(2h,2h)$节点值所作的线性插值相同；I_{2h}^h的插值精度阶是多少？

19. 使用下述三次插值：设(x,y)是Ω_l中的网格节点，$(x\pm h,y)$和$(x\pm 3h,y)$是Ω_{l-1}中的网格节点，在(x,y)处以及$(x\pm\nu h,y\pm\mu h)$(ν、μ为偶数)处的x方向插值公式形如

$$
\begin{aligned}
v_l(x,y)=(pv_{l-1})(x,y)=&-\frac{1}{16}v_{l-1}(x-3h,y)\\
&+\frac{9}{16}v_{l-1}(x-h,y)+\frac{9}{16}v_{l-1}(x+h,y)\\
&-\frac{1}{16}v_{l-1}(x+3h,y);
\end{aligned}
$$

在$(x\pm\nu h,y\pm\mu h)$(ν、μ为奇数)处，则类似地作出y方向上的插值。若将这种三次插值写成局部离散格式，则有

$$
\begin{pmatrix}
1/16^2 & 0 & -9/16^2 & -1/16 & -9/16^2 & 0 & 1/16^2\\
0 & 0 & 0 & 0 & 0 & 0 & 0\\
-9/16^2 & 0 & 81/16^2 & 9/16 & 81/16^2 & 0 & -9/16^2\\
-1/16 & 0 & 9/16 & 1 & 9/16 & 0 & -1/16\\
-9/16^2 & 0 & 81/16^2 & 9/16 & 81/16^2 & 0 & -9/16^2\\
0 & 0 & 0 & 0 & 0 & 0 & 0\\
1/16^2 & 0 & -9/16^2 & -1/16 & -9/16^2 & 0 & 1/16^2
\end{pmatrix}.
$$

试证明此插值对于任一双三次函数$\sum_{\nu=0}^{3}\sum_{\mu=0}^{3}a_{\nu,\mu}x^\nu y^\mu$是精确的。

20. 试构造对应于矩形分划在内节点处适用的分片双二次插值局部离散格式。

21. 编制使用5点离散格式求解Poisson方程第一边值问题的MG方法(V循环或W循环)程序，其中定解区域$\Omega=[0,1]\times[0,1]$将被分为$2^m\times 2^m$等分，相应的网格层记为Ω_m，光滑方法一律采用GS方法。试就一个边值问题的例子，计算并考察下

列问题。

(1) 用相同的初值，用$\|u_m^k - u\| = 0.1h_m^2$作为迭代精度控制，分别在Ω_m($m = 1, 2, \cdots, 8$)上计算出数值解$u_1, u_2, \cdots, u_8$。

(2) 计算u_m时，记录比较每个MG循环步的收敛因子。

(3) 根据(2)计算出MG循环步的平均收敛因子$\bar{\rho}_m$。

(4) 比较$\bar{\rho}_2, \bar{\rho}_3, \cdots, \bar{\rho}_8$，估计$\bar{\rho}_m$($m \to \infty$)的$h$无关特征。

22. 试构造在$\bar{\Omega} = [0, 1] \times [0, 1]$作RB分划情形下的多层网格方法。

23. 试证明完全多层网格方法的h无关收敛性。

24. 试证明在有限元多层网格方法中，下式成立(参见§5.2)：

$$L_{l-1} = I_l^{l-1} L_l I_{l-1}^l.$$

25. 试指出下述"逼近特性"证明中的问题：

$$\|L_l^{-1} - I_{l-1}^l L_{l-1}^{-1} I_l^{l-1}\| = \|L_l^{-1} - I_{l-1}^l (I_l^{l-1} L_l I_{l-1}^l)^{-1} I_l^{l-1}\|$$
$$= \|L_l^{-1} - I_{l-1}^l (I_l^{l-1} L_l I_{l-1}^l)^{-1} I_l^{l-1}\|$$
$$= \|L_l^{-1} - I_{l-1}^l (I_{l-1}^l)^{-1} L_l^{-1} (I_l^{l-1})^{-1} I_l^{l-1}\| = 0 \leqslant ch_l^\alpha.$$

26. 试编制解矩形域上Poisson方程的逐层子空间迭代程序，并就具体例子计算考察下列问题：

(1) 在最细网格层上计算出节点值u_l、节点导数值du_l和超收敛导数值Du_l，并计算出它们的外推值$\tilde{u}_l$、$d\tilde{u}_l$和$D\tilde{u}_l$；

(2) 作出u_l、$\tilde{u}_l$、du_l、$d\tilde{u}_l$、Du_l、$D\tilde{u}_l$的事后误差估计，且将其与事先误差估计比较。

27. 已知细网格层上的离散方程$L_l u_l = f_l$和任一近似值u_l^j，可用逐层子空间迭代法解残差方程$L_l w_l = d_l^j = f_l - L_l u_l^j$的近似解$w_l^j$，从而获得校正后的新近似值$u_l^{j+1} = u_l^j + w_l^j$。试将这个新的迭代与MG方法比较。

28. 已知局部离散格式

$$[S_k] = \begin{bmatrix} & -1 & \\ -1 & 4 & -1 \\ & -1 & \end{bmatrix},$$

$$[\bar{S}_k]=\begin{pmatrix} a & b & a \\ b & c & b \\ a & b & a \end{pmatrix},\qquad [\tilde{S}_k]=\begin{pmatrix} & & * & & \\ & * & 0 & * & \\ * & 0 & * & 0 & * \\ & * & 0 & * & \\ & & * & & \end{pmatrix},$$

其中 * 和 0 分别表示非零元素和数字零。试就关系

$$[\bar{S}_k][S_k]=[\tilde{S}_k]$$

利用待定系数法确定$[\bar{S}_k]$和$[\tilde{S}_k]$中的非零元素。

29. 对 5 点格式

$$[S]_1 u(x)=-\Delta u h^2-\frac{h^4}{12}\Delta^2 u+O(h^6),\qquad x\in\Omega_h,$$

两边同乘局部离散格式$[S]_3 \cdot [S]_2$，就有格式

$$[S]_4 u(x)=128h^2 f+16h^4\Delta f-\frac{32}{8}h^4\Delta^2 u+O(h^6),\qquad x\in\Omega_{2h},$$

其中，$\Delta^2 u=\frac{\partial^4 u}{\partial x^4}+\frac{\partial^2 u}{\partial y^4}$，

$$[S]_1=\begin{pmatrix} & -1 & \\ -1 & 4 & -1 \\ & -1 & \end{pmatrix}_h,\qquad [S]_2=\begin{pmatrix} & 1 & \\ 1 & 4 & 1 \\ & 1 & \end{pmatrix}_h,$$

$$[S]_3=\begin{pmatrix} -1 & 2 & -1 \\ 2 & 12 & 2 \\ -1 & 2 & -1 \end{pmatrix}_h,\qquad [S]_4=\begin{pmatrix} & & 1 & & \\ & -2 & -32 & -2 & \\ 1 & -32 & 132 & -32 & 1 \\ & -2 & -32 & -2 & \\ & & 1 & & \end{pmatrix}_{2h}$$

(1) 试验证第二个格式的正确性，

(2) 试用这两种格式外推，导出在 $x\in\Omega_h\cap\Omega_{2h}$ 处的新的高精度格式，并指出新格式的截断误差。

30. Kirchhoff 变换可将某类非线性椭圆型方程化为线性方程。例如，若令 $v=v(u)$，$\frac{dv}{du}=k(u)$，则可将非线性方程

$$\nabla[k(u)\nabla u]=0$$

化为关于 v 的线性方程

$$\Delta v=0.$$

试详细推导这一过程，并讨论 Kirchhoff 变换在数值求解此类非线性椭圆边值问题中的应用可能性。

参考文献

[1] 武汉大学数学教研室译，椭圆型方程差分方法，科学出版社，1984.

[2] 蒋叔豪、孙庆新编，偏微分方程数值解，浙江大学出版社，1985.

[3] 蔡大用、施妙根译，实用迭代法，清华大学出版社，1984.

[4] 李荣华、冯果忱编，微分方程数值解法，人民教育出版社，1981.

[5] M.R.Hestens *and* E.L.Stiefel, *Methods of Conjugate gradients for solving linear systems*, Nat. Bur. Std. J. Res., Vol. 49, pp. 409-436, 1952.

[6] P. Concus, *A generalized conjugate gradient method for the numerical solution of elliptic partial differential equations*, Sparse Matrix Computation (J. R. Bunch *and* D. J. Rose eds.), pp. 309-332, Academic Press, New York, 1976.

[7] D. J. Rose *and* R. A. Willoughby eds., *Sparse Matrices and Their Applications*, Plenum Press, New York, 1972.

[8] J. A. Meyerink *and* H. A. v. d. Vorst, *An iterative solution method for linear systems of which the coefficient matrix is a symmetric M-matrix*, Math. Comp., Vol. 31, pp148-162,1977.

[9] F. S. Beckman, *The solution of linear equations by the conjugate gradient method*, Mathematical Methods for Digital Computers (A. Ralsten *and* H. S. Wilf eds.), Vol. I, pp. 62-72, 1960.

[10] A. Brandt, *Multi level adaptive technique (MLAT) for fast numerical solution to boundary value Problems*, Notes in Physics, Vol. 18, pp. 82-89, Springer-Verlage Berline, 1973.

[11] A. Brandt, *Multi level adptive solutians to boundary value*

problems, Math. of Computation, Vol. 31, Num. 138, pp.33—390, 1977.

[12] W. Hackbusch *and* U. Trottenberg *eds.*, *Lecture Notes in Mathematics* 960, Springer-Verlage, Berline, 1982.

[13] W.哈克布思著，林群译，多重网格方法，科学出版社，1988.

[14] R. E. Bank, *An optimal order process for solving finite element equations*, Math. of Computation, Vol. 36, Num. 153, pp. 35—51, 1981.

[15] K. Brand, *Multigrid Bibliography*, Edition 3. Nr. 59, Arbeitapapier der GMD, 1983.

[16] K. Becker, COMFLO *ein Experimentier-Programm zur Mehrgitten Behandlung subsoni-scher Potentialströmungen um Tragflächenprofile*, Preprint *no*, 604., Sonderforschungs-bereich 72 *in* Uni. Bonn.

[17] 黄鸿慈、刘贵银，解椭圆型边值问题的逐步加密法,计算数学，1978.

[18] 黄鸿慈等，矩形域低光滑解的渐近展开，计算数学，1986.

[19] 朱起定，有限元超收敛理论，湖南科技出版社，1989.

[20] Lin Qun、Lu Tao *and* Shen Shumin, *Maximum norm estimate, extrapolation and optimal points of stresses for the Finite Element Methods on the strongly regular triangulation*, J. Comp, Math., pp 376-383, 4, 1983.

[21] Lin Qun *and* Wang Junping, *Some expansion of the finite element approximation*, Cheng du Branch *of* Academia Sinica, 1984.

[22] Lin Qun *and* J. Q. Liu, *Extrapolation method for Fredholm integral equation with non-smooth kernels*, Numer. Math., pp 459-464, 35, 1980.

[23] 陈传淼，矩形双二次元的外推法，湘潭大学自然科学学报，pp.1—6, 3,1987.

[24] 陈传淼，Galerkin解的外推法，湘潭大学自然科学学报，4，1981.

[25] 陈传淼，拟线性椭圆边值问题有限元解的渐近展式及其外推，东北数学杂志，pp.228—236，2，1986.

[26] 徐长发，关于 Dirichlet 边值最小面积问题的有限元解法，华中工学院学报，6，1981.

[27] 徐长发，有限元多层网格方法的代数解释，华中工学院学报,2,1985.

[28] 徐长发，有限元多层网格方法（V 循环）实用参数选择的研究，固体力学学报，3，1986.

[29] 徐长发，有限元方程 Jacobi 迭代的收敛性分析，华中工学院学报，2，1985.

[30] 徐长发，一种求解有限元方程的新方法，华中工学院学报，2，1985.

[31] 徐长发等，逐层子空间迭代收敛性分析，华中工学院学报，1，1987.

[32] 徐长发等，一类非线性抛物型方程有限元解的误差估计，华中工学院学报，5，1987.

[33] 徐长发等，一类非线性双曲型方程有限元解的误差估计，华中工学院学报，6，1987.

[34] 徐长发等，有限元方法与差分方法的某些统一性，4，1989.

[35] 徐长发，解有限元方程的逐层分裂迭代法，应用数学，1，1988.

[36] Q. Lin *and* T. Lu, *Correction and splitting procedures for solving* PDE. Math. Sci., Reseach Report. IMS-9, 1985.

[37] H. Munz. *Uniform expansions for a class of finite difference schemes for elliptic boundary value problems*, Math. Comp., pp.155-170, 36(1981).

[38] R. P. Fedorenko, *A relaxation method for solving elliptic difference equations*, USSR Comp. Math. Math. Phys., pp.1092-1096, 1, 1962.

[39] J. Schröder *und* U. Trottenbery, *Reduktionsverfahren für Differenzengleichungen bei Randwertaufgaben I*, Numer. Math., pp.37-68, 22, 1973.

[40] J. Schröder、U. Trottenberg *und* H. Reutersberg, *Reduktionsverfahren für Differenzengleichungen bei Randwertaufgaben* Ⅱ, Numer. Math., pp.429-453, 26,1976.

[41] K. Böhmer, *Discrete Newton methods and iterated defect corrections*, Numer. Math. 37, 167-192, 1981.

[42] L. Fox, *Numerical solution of ordinary and partial differe*

ntial equations, Oxford Pergamon Press, 1962.

[43] H. J. Stetter, *The defect correction principle and discretization methods*, Numer. Math. 29, 425-443, 1978.

[44] K. Böhmer, *A defect correction method for functional equations, in Approximation Theory*, R. Schaback *and* K. Scherer, eds., Lecture Nates in Mathematics 556, Springer-Verlag, Berlin, pp. 165-170, 27, 1977.

[45] 郭本瑜著，偏微分方程的差分方法，科学出版社，1988。